Historic, Archive Document

Do not assume content reflects current scientific knowledge, policies, or practices.

Some Notes On

LEAST SQUARES

by

W. Edwards Deming

The Graduate School

The Department of Agriculture

Washington

Price $1.50

Some Notes On

LEAST SQUARES

by

W. Edwards Deming, Ph. D.

Assistant Director of The Graduate School, and
Physicist, Bureau of Chemistry & Soils,
The Department of Agriculture
Lecturer in Mathematics, National Bureau of Standards
Washington

The Graduate School

The Department of Agriculture

Washington

Price $1.50

Mimeographed in the U. S. A.

PREFACE

A bit of history will be helpful to anyone who may be wondering just how this brochure happens to be in existence. Beginning in 1932 the author has given annually a course in statistics at the Department of Agriculture Graduate School. The course consists of a survey of classical probability and an introduction to modern theoretical statistics, with some philosophy mixed in with it in an attempt to separate the mathematical developments from observations on the physical behavior of nature. The last six weeks of the course are devoted to the study of least squares.

It is the lecture notes for this part of the course for 1937 that now appear. When the good and harm of their public appearance have been separated and weighed, whatever excess there be in the positive direction is to be credited to my friend Miss Besse B. Day of the Forest Service. In December 1937 she asked permission to reproduce my manuscript for the use of her colleagues in Washington and in the field, some of whom it has been my privilege to have worked with, in and out of classes. I pleaded for a few weeks in which to put the notes into better order. The task seemed hopeless; they had grown up not with the notion that they were to form a text, complete in itself, but simply to serve in the last part of the course described earlier. But after some thought in the matter, it became clear that there is a widespread need for just this type of material; accordingly Miss Day's original plan has been carried out. The assumption is that the reader knows when he needs to use least squares and when he should not, that is, when the question can be answered at all. In the course of study in which these notes have grown up, such things are discussed before least squares is commenced; the material presented here may therefore be regarded purely as an exposition on the principle and method. Naturally I have been partial to my own experience, which has embraced consultation on a wide variety of problems, mostly in the Department of Agriculture and the National Bureau of Standards. I have tried to stress points that have seemed inadequately covered or incorrectly pictured by previous writers. It is presumed that teachers using this as a text will heed the references for supplementary reading, and will also illustrate the theoretical exercises with numerical examples.

During the years 1922-23 it was my good fortune to have had as a teacher Professor Charles A. Hutchinson of the University of Colorado; later, Professor Horace S. Uhler of Yale, to whose paper of 1923 I owe much of the inspiration for whatever is new in the present work. More recently I have had the privilege of either studying or working with R. A. Fisher, J. Neyman, and Egon S. Pearson in London, and Walter A. Shewhart of New York, thus to have been exposed to several different views of the ever lively question, how did we learn what we think we know, and what do we really know?

Most of the computations in the following pages were performed by my wife, Lola S. Deming. The typing is the work of Miss Eleanor Wood of the Forest Service, whose skill and patience in writing mathematics will be appreciated by the readers as much as it has been by the author. Equally valuable has been the assistance in checking and proof-reading rendered by Miss Day and Miss Marion M. Sandomire, also of the Forest Service; a number of blunders that they uncovered in the manuscript would have been to me a perennial source of embarrassment. A final and expert perusal of the stencils by Lee Garby (Mrs. C. D. G.), of the Bureau of Chemistry and Soils, has been no less contributory. Lastly, it is a pleasure to record an appreciation of my counselor and chief, Dr. C. H. Kunsman, on account of his never failing encouragement and enthusiasm for mathematics during my eleven years in the government service.

As the labor of putting these notes into print is being concluded, I am conscious of many failings, both by omission and commission; yet the words of Thomas Simpson in the preface of the first edition of his Treatise of Algebra (London, 1745) give an accurate expression of my own thoughts:

> As to the Motives that first gave Birth to the ensuing Work, they were not so much any extravagant Hopes I could form to myself of greatly extending the Subject by the Addition of a large Variety of new Improvements, though, the Reader will find many Things here more than he has yet seen, and perhaps than he expects, as an earnest Desire of seeing a Subject of so much Importance established on a clear and rational Foundation, ...

W. E. D.

Washington
May 1938

TABLE OF CONTENTS

III. APPLICATION TO CURVE FITTING

Some Notes On

L E A S T   S Q U A R E S

" ... quamobrem systema maxime probabile valorum pro quantitatibus p, q, r, s etc. id erit, ubi aggregatum hhvv + h'h'v'v' + h"h"v"v" + etc. i.e. ubi summa quadratorum differentiarum inter valores revera observatos et computatos per numeros qui praecisionis gradum metiuntur multiplicatarum fit minimum." Karl Friedrich Gauss, Theoria Motus Corporum Coelestium, (Hamburg, 1809). Art. 179.

1. Preliminary; some fragments in the history of least squares. It is helpful in the study of any subject to keep an eye on its history. But as Max Nordau* has pointed out, history is elusive; there are errors of observation in history of a grosser kind than in the laboratory; moreover, there is always a personal judgment as to just what events in the past really contributed to later ones. In what follows, the numerous references to Gauss (1777-1855) will leave no doubt in the reader's mind that in my own judgment he accomplished most of what we think we know now about least squares.

But of course there were others. A hundred years before Gauss' time, De Moivre** originated the normal curve (1733), which was tabled by Kramp+ in 1799. Multitudes since the time of Gauss have helped to clarify the basic principles involved by looking askance at theoretical statistics and in particular anything developed from the normal curve++.

---

* Max Nordau, The Interpretation of History (Transl., Rebman, London, 1910).

** Karl Pearson, Historical note, Biometrika 16, 402-404, 1924. A facsimile of the famous seven page pamphlet by De Moivre appears in an article by R. C. Archibald in Isis 8, October 1926.

\+ Kramp, Analyse des réfractions (Strasbourg, 1799). Some historical notes regarding this book will appear in an article by Mason DuPré in Isis, 1938, possibly 1939.

++ Gauss himself cast a critical eye at it, but satisfied himself that deductions from it must be useful, witness the following quotation from his Theoria Motus Art. 178: "Functio modo eruta omni quidem

There were Laplace and Legendre whom Gauss mentioned; Encke, Bessel, Airy, and Helmert, and I suppose many others. In 1879 came a neglected paper by Kummell (vide infra) treating the problem of curve fitting with great clarity. In 1877 Mansfield Merriman* made up a list of all the writings on least squares that had come to his notice--408 books and papers altogether. He hardly hoped that the list could be complete, but it is of great value. Of course he missed De Moivre, whose work was lost until Karl Pearson uncovered it in 1924.

Gauss, Bessel, Airy, and Helmert were astronomers, and in most of the problems of adjustment that they met, the conditions imposed on the adjusted observations were geometrical, not involving adjustable parameters, and the result was that many questions regarding curve fitting were left for a later time. Kummell** in 1879 seems to have filled in a good share of what was missing. Unfortunately, his work fell on poor soil; the journal that he wrote in was obscure and his paper has had very little influence on textbooks or general usage. He set down the condition equations for the problem of curve fitting when more than one coordinate is subject to error, and he remarked that the situation was hopeless except in some of the simplest cases, one of which he proceeded to treat--the straight line, for which he obtained a solution under the supposition that the weights of the x and y coordinates are in a constant ratio for all points (see Ex. 6 of section 19). He showed how the weights of the estimated parameters a, b, c could be obtained from the normal equations along with the sum of the squares of the weighted residuals, if ever the problem is not too complex to permit the equations to be set up. The use of the Lagrange multipliers, had he tried them, would have provided the general solution that he was looking for (section 9 et seq.).

Karl Pearson+ in 1901 wrote "On lines and planes of closest fit to systems of points in space". Therein he obtained the best

---

rigore errorum probabilitates exprimere certo non potest: quum enim errores possibiles semper limitibus certis coërceantur, errorum maiorum probabilitas semper evadere deberet = 0, dum formula nostra semper valorem finitum exhibet. Attamen hic defectus, quo omnis functio analytica natura sua laborare debet, ad omnes usus practicos nullius momenti est, quum valor functionis nostrae tam rapide decrescat, quamprimum $h\Delta$ valorem considerabilem acquisivit, ut tuto ipsi 0 aequivalens censeri possit. Praeterea ipsos errorum limites absoluto rigore assignare, rei natura numquam permittet". (His $h\Delta$ is proportional to the more usual $x/\sqrt{2}\sigma$).

* Mansfield Merriman, Connecticut Academy of Arts and Sciences 4, 151-232, 1877-82.

** Charles H. Kummell, The Analyst (Des Moines) 6, 97-105, 1879.

\+ Karl Pearson, Phil. Mag. (London) 2, 559-572, 1901.

and worst fitting lines and planes, together with ingenious and now well known formulas for the mean square residual, the implicit premise being that all observations have the same weight (cf. Ex. 9 in section 19). Pearson did not make the assumption of small residuals; his results are exact. He evidently had not heard of Kummell's paper.

Papers by Stewart* (1920) and Uhler**(1923) should be consulted for a careful explanation of the principle of least squares, especially in regard to curve fitting. In 1921 Corrado Gini+ found the best and worst fitting lines for points whose x and y coordinates are in a constant ratio from point to point, the main results having already been reached by Kummell. In the same issue of Metron in which Gini wrote, Lowell J. Reed++ described a clever mechanical device for fitting a line when the x and y coordinates have equal weights everywhere; all that is needed is a fine smooth steel rod and a number of uniform elastic bands, one for each point.

Going back to Gauss, one can say that he not only laid the theoretical foundations for least squares but he also gave careful attention to methods of calculation, and he put them into practice. He developed an abridged solution of the normal equations, and showed how to find the weights of functions of the adjusted values, and the sum of the weighted squares of the residuals, right along with the solution. The Gauss solution is found in most texts; in particular Comstock's° should be mentioned. Doolittle's method, and many modifications of it, are all based on Gauss' plan, which took full advantage of the symmetry of the normal equations. Naturally, modern electric calculating machines have brought about changes in routine here and there.

In the present day, least squares can not be dissociated from the statistical researches that have been made since the writings of Gauss, Encke, and Helmert appeared. Most of today's students of statistics are familiar with what has been done in that line in our own generation, and no description of it need be made here; instead, some of it will be woven into the material that is to follow.

---

* R. Meldrum Stewart, Phil. Mag. (London) 40, 217-227, 1920.

** Horace S. Uhler, J. Optical Society 7, 1043-1066, 1923.

\+ Corrado Gini, Metron (Rome) 1, No. 3, 63-82, 1921.

++ Lowell J. Reed, Metron 1, No. 3, 54-61, 1921.

° George C. Comstock, Method of Least Squares, (Ginn, 1890).

On the part of many recent writers in statistics it can be said that an occasional lack of familiarity with the memoirs of Gauss and his contemporaries has caused no little confusion in the minds of students. More than once a profound "discovery" in modern statistics has turned out to be something already known under other forms to students with classical training. At the same time, many modern writers in least squares, especially of textbooks, have ignored what was being developed in statistical studies, and have, moreover, developed the subject of least squares from an incoherent perspective of the adjustment of observations as it occurs under the varieties of conditions that may be imposed. All this has retarded progress. The pages that follow may possibly be a step--they can not be more--toward unifying some closely related topics in least squares and statistics. A knowledge of least squares is, to my mind, requisite even if one prefers to use other methods of curve fitting.

I. SOME SIMPLE PROBLEMS IN CURVE FITTING

2. The principle of least squares. Before going into the general problem of the adjustment of observations by least squares, it will be helpful to look at some very simple applications in curve fitting. It is a fact that the simple ones afford nearly as much opportunity for thought in the field of statistical inference as the more complicated ones do. In all of them the principle of least squares requires the minimizing of χ^2, where

$$\chi^2 = (1/\sigma^2) \sum w \text{ res}^2 \tag{1}$$

the summation of the weighted squares of residuals to be taken over all observations, σ being the r. m. s. error of observations of unit weight. This is the equivalent of Gauss' pronouncement quoted at the opening. In curve fitting, both x and y observations may be subject to error, and the summation of the weighted squares of residuals will then include both coordinates, and will occasionally be written explicitly as $\Sigma(w_x V_x^2 + w_y V_y^2)$. For an explanation of the symbols V_x and V_y see Fig. 9 on page 83. For convenience, the symbol ϕ^2 will sometimes be used to designate the summation of the weighted squares of residuals, as in Eqs. 3, 8, 30, and elsewhere; and we may then simply say that

$$\chi^2 = \phi^2/\sigma^2 \tag{1a}$$

For an interpretation of χ^2 see the exercises in section 5.

Now since σ is a constant for one particular problem, χ^2 is a minimum when ϕ^2 is a minimum; hence we may think of least squares not only as the minimizing of χ^2 but also of ϕ^2, i. e., of the sum of the weighted squares of the residuals. Least squares may also be considered the minimizing of the estimate $\sigma(\text{ext})$, as will be explained in section 6c.

Another way of looking at the problem is to say that the principle of least squares is the <u>maximizing</u> of $P(\chi)$, and that we seek the solution that gives the greatest probability on the chi-test.

The <u>principle</u> of least squares has been stated. The <u>method</u> of least squares is a rule or set of rules for going about the actual computation.

3. The simplest example of curve fitting is the single sample of n observations of equal weight on an unknown quantity α. Here the curve to be fitted is the simplest of all curves

$$x = a \qquad (2)$$

containing just one "parameter", namely a, which is to be estimated by least squares. Here in this simple case where the weights* are all unity,

$$\chi^2 = (1/\sigma^2)\,\Sigma\,\text{res}^2$$
$$= (1/\sigma^2)\,\Sigma\,(x_i - a)^2 \qquad (3)$$

The y coordinates of the points are merely the ordinal numbers of the observations, and hence are without error, so only x-residuals appear in the summation. Now χ^2 is a minimum when the numerator

$$\phi^2 = \Sigma(x_i - a)^2 \qquad (4)$$

is a minimum. The observations, having once been made, can not be changed, hence the only variable in Eq. 4 is a. By giving various

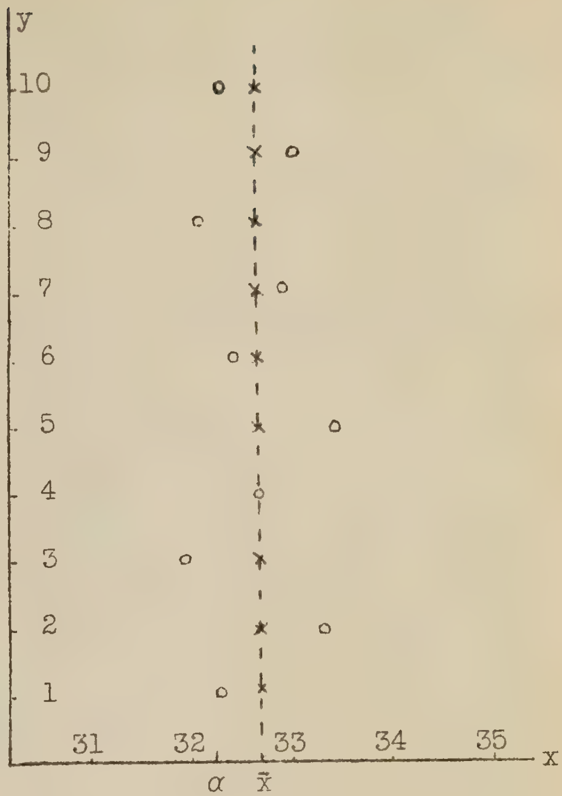

Fig. 1. Ten observations

$x_1,\ x_2,\ \ldots\ ,\ x_{10}$

of equal precision (equal weight) are made on an unknown magnitude α. The true points are connected by the simple relation $x = \alpha$, hence $x = a$ is the curve to be fitted. The least squares value of a turns out to be $\bar{x}$, the mean of the ten observations. $x = \alpha$ is the "true curve", and $x = \bar{x}$ is the "calculated curve". The calculated points are shown by the crosses; they all lie on the calculated curve.

---

* For a discussion on weights see section 5.

values to a, ϕ^2 is made to take on various values, and the minimum will occur when the derivative*

$$d\phi^2/da = -2\sum(x_i - a) \qquad (5)$$

vanishes, that is to say, when

$$\sum(x_i - a) = 0 \quad \text{or} \quad \sum x_i - na = 0$$

The least squares value of a is accordingly

$$a = (1/n)\sum x_i \equiv \bar{x} \qquad (6)$$

So the vertical line $x = \bar{x}$ in Fig. 1 gives the smallest possible value to the sum of the squares of the residuals, and hence to χ^2. The line $x = \bar{x}$ is the "calculated curve"; it is the "curve" x = a fitted by least squares. On this line lie all the "calculated points", these being the least squares estimates of the observed points. In this very simple problem, the calculated points all have the same abscissa, namely $\bar{x}$. (Compare Fig. 1 with Fig. 9, p. 83).

The goodness of fit may be judged by the value of $\chi^2/(n-1)$ compared with σ^2, in other words by $P(\chi)$ for n-1 degrees of freedom. This subject will be touched again in section 6c and elsewhere.

Note that the value of σ is not required for the application of least squares, since whatever σ is, χ^2 is a minimum when ϕ^2 is. σ did not occur in Eq. 5. σ is required, nevertheless, for the use of the chi-test for the goodness of fit obtained.

Note also that if s denotes the S. D. of the n measurements, then $\chi^2 = ns^2/\sigma^2$, and the minimized value of ϕ^2 is ns^2. For a

---

* $d\phi^2$ is to be interpreted as the differential of ϕ^2, not as $(d\phi)^2$. $d\phi^2/da$ is the derivative of ϕ^2 with respect to a.

new sample of n observations there will be a new mean $\bar{x}$, a new line, and a new χ^2.

4. The same problem with unequal weights. (a) The solution for the parameter a. Suppose that the n observations $x_1, x_2, \ldots, x_n$ have weights $w_1, w_2, \ldots, w_n$, perhaps not all equal; then

$$\chi^2 = (1/\sigma^2) \Sigma w_i (x_i - a)^2 \tag{7}$$

the i-th residual being, as before, $x_i - a$. We are now to make

$$\phi^2 = \Sigma w_i (x_i - a)^2 \tag{8}$$

a minimum with respect to a. The derivative

$$d\phi^2/da = -2 \Sigma w_i (x_i - a)$$

is to be set equal to zero, giving

$$\Sigma w_i (x_i - a) = 0$$

i.e.
$$a \Sigma w_i = \Sigma w_i x_i \tag{9}$$

or
$$a = \Sigma w_i x_i / \Sigma w_i \equiv \bar{x} \tag{10}$$

where $\bar{x}$ is now the "weighted mean" of the n observations. In the event that $w_1 = w_2 = \ldots = w_n$, this result reduces to the previous value of a in Eq. 6; in other words, the problem of the preceding section was a special case of this one.

The minimized value of ϕ^2 is here

$$\phi^2 = \Sigma w_i (x_i - \bar{x})^2 = \Sigma w_i x_i^2 - \bar{x}^2 \Sigma w_i \tag{11}$$

(b) The solution in tabular form. In sections 16 and 17 we shall see a systematic procedure for the solution of normal equations and for calculating the "reciprocal matrix", in which are found the variance and product variance* coefficients; also we shall see the minimized value of ϕ^2 calculated right along with the solution of the normal equations. In simple problems like the one just considered, there is only one normal equation (Eq. 9) and it is of course very easily solved (see Eq. 10). Nevertheless it is interesting to see how the routine process that is to be learned later applies here; let us therefore set up the following tabulation, and perform the steps indicated, not worrying why. When more complicated problems are followed through, as in sections 16-20, it will be helpful to recall that the same procedure led to known results back here in simpler cases.

No.	a	=	1	C
1	Σw		Σwx	1
2			Σwx^2	0

If the top row is divided through by Σw (found under a) we find

3	1	$\Sigma wx/\Sigma w$	$1/\Sigma w$

Interpreted this means that

$$a = \Sigma wx/ \Sigma w$$

as already seen in Eq. 10. Moreover, in the C column of No. 3 we find $1/\Sigma w$, which can be considered the one and only term in the reciprocal

---

* I use the term product variance rather than covariance, following Dr. A. C. Aitken of Edinburgh.

matrix, that term being the reciprocal of the weight of a-- in other words, the variance coefficient of a, which is to say that

$$\sigma_a^2 = \sigma^2/w_a = \sigma^2/\sum w \qquad (12)$$

This equation will be understood better after the discussion on weights has been read (next section).

Now let No. I be multiplied through by $-\sum wx/\sum w$ and added to No. 2; we find in the "1" column (the other columns are disregarded)

$$\sum wx^2 - (\sum wx)^2/\sum w, \text{ or } \sum wx^2 - \bar{x}^2 \sum w$$

and this is none other than the minimum value of ϕ^2, which has come forth without the intermediate step of computing each residual and squaring it. These results will have a fresh significance after sections 16 ff have been studied; see in particular exercise 3c of section 17 (p.109), and applications in section 19.

5. A digression to define weights. By definition, the weight w_f of the function f is inversely proportional to the variance σ_f^2 of f. That is to say, $1/w_f$ is the variance coefficient of f. In symbols,

$$w_f = \sigma^2/\sigma_f^2 \quad \text{or} \quad \sigma_f^2 = (1/w_f)\sigma^2 \qquad (13)$$

σ^2 is simply a proportionality factor, and is evidently the variance of a function of unit weight. If σ^2 be arbitrarily doubled, and w_f also doubled, σ_f^2 is unaffected in value.

For example, let f be $\bar{x}$, the mean of the n observations x_1, x_2, ..., x_n, which are random variates taken from a universe of S. D. σ, hence each of unit weight. Then, since the variance of $\bar{x}$ is σ^2/n, Eq. 13 tells us that

$$w_{\bar{x}} = \sigma^2/(\sigma^2/n) = n \quad \text{or} \quad \sigma_{\bar{x}}^2 = (1/n)\sigma^2 \qquad (14)$$

whence we see that n is the weight, and 1/n the variance coefficient, of $\bar{x}$. Or, if the n original observations were each of weight w instead of unity (as we could as well say, since weights are relative and not absolute, depending as they do on the arbitrary factor σ^2), then the variance of single observations would be σ^2/w, and the variance of $\bar{x}$ would be one nth as much, so in this case Eq. 13 gives

$$w_{\bar{x}} = \sigma^2/(\sigma^2/nw) = nw \quad \text{or} \quad \sigma_{\bar{x}}^2 = (1/nw)\sigma^2 \quad (15)$$

saying that nw is now the weight, and 1/nw the variance coefficient, of $\bar{x}$. So, as before, the weight of $\bar{x}$ is just n times the weight of a single observation.

As E. B. Wilson* stated it: "The primal conception of a weight is that of a repeated observation." In Fisher's terminology, the

---

* E. B. Wilson and Ruth R. Puffer, "Least squares and laws of population growth," Proc. Amer. Acad. Arts and Sci. (Boston) 68, No. 9, August 1933.

mean $\bar{x}$ of n normal observations contains n times as much information as a single observation.

Concerning two functions f_1 and f_2, it can be said at once from Eq. 13 that

$$w_1 : w_2 = \sigma_2{}^2 : \sigma_1{}^2 \tag{16}$$

which says that the weights of two functions are inversely proportional to their variances.

Exercise 1. Given $\chi^2 = (1/\sigma^2) \sum w V^2$ as in Eq. 1, show that χ^2 may be written

$$\chi^2 = \sum \left(\frac{V}{\sigma/\sqrt{w}}\right)^2$$

that is, χ^2 is the sum of the squares of the residuals, each residual being measured in units of the S.E. $\sigma/\sqrt{w}$ of the corresponding observation of weight w (compare with Eq. 20, next section). In other words, χ^2 is the sum of the squares of the standardized residuals. χ^2 is therefore independent of the units used in measurement; a change from feet to inches or centimeters, or from pounds to ounces or grams, changes the residuals, but not the standardized residuals, nor χ^2.

Exercise 2. When both x and y observations are subject to error, one may wish to designate the summation explicitly as

$$\chi^2 = (1/\sigma^2) \sum (w_x V_x{}^2 + w_y V_y{}^2)$$

as has already been indicated in section 2. Show that this may be written

$$\chi^2 = \sum \left\{\left(\frac{V_x}{\sigma/\sqrt{w_x}}\right)^2 + \left(\frac{V_y}{\sigma/\sqrt{w_y}}\right)^2\right\}$$

which again says that χ^2 is the sum of the squares of all the residuals, each one being measured in units of the S.E. of the corresponding observation on the x or y coordinate (see the figure on page 83). So χ^2 is,

as before, the sum of the squares of the standardized residuals. The remarks in the preceding exercise hold.

Exercise 3. ϕ^2, or the sum of the weighted squares of the residuals is, like χ^2, also invariant to changes in units (as from pounds to ounces, etc.). But ϕ^2 is dependent on the arbitrary choice of σ, whereas χ^2 is not. One weight in the whole set is arbitrary, and the others are related to it through Eq. 13; fixing this one weight is equivalent to fixing σ. ϕ^2 can be doubled by doubling all the weights, but this has no effect on χ^2 because σ^2 would also be doubled. The least squares solution for a (and other parameters, if any, as in more complicated problems) is independent of σ^2; the parameter or parameters that minimize ϕ^2 for one set of weights will also minimize it if all the weights are doubled.

Note: For another interpretation of ϕ^2 in curve fitting, see Ex. 3 of section 15, page 97.

For other exercises in weights, see page 30.

6. Several samples, all on the same unknown. (a) All observations have the same precision. Let us suppose that n observations of equal weight (equal precision) and all on the same unknown, as for example those of section 3, are arbitrarily subdivided into m samples of $n_1, n_2, \ldots, n_m$ observations. We shall say that

X_1 is the mean of n_1 single observations
X_2 " " " " n_2 " "
⋮ ⋮
X_m " " " " n_m " "

Now if single observations have unit weight, then it will follow from Eqs. 14 or 15 that the weights of the m means are

$$w_1 = n_1, \; w_2 = n_2, \; \ldots, \; w_m = n_m$$

We may now consider the m sample means to be m observations of weights respectively $n_1, n_2, \ldots, n_m$, to which the results of section 4 apply. The value of a that minimizes χ^2 is then by Eq. 10,

$$X = \frac{\Sigma wX}{\Sigma w} = \frac{n_1X_1 + n_2X_2 + \ldots + n_mX_m}{n_1 + n_2 + \ldots + n_m} \tag{17}$$

This X is the "weighted mean" of the m samples; it is simply the arithmetic mean of the $n_1 + n_2 + \ldots + n_m$ single observations since they all are of equal weight. Deviations (V_i) reckoned from X make

$$\chi^2 = (1/\sigma^2)\, \Sigma wV^2$$

a minimum, and it is noteworthy that, at the same time, $\Sigma V = 0$.

It should be kept in mind that X is independent of just how the n observations are subdivided, but χ^2 is not. Moreover, both X and χ^2 will vary from one set of observations to another.

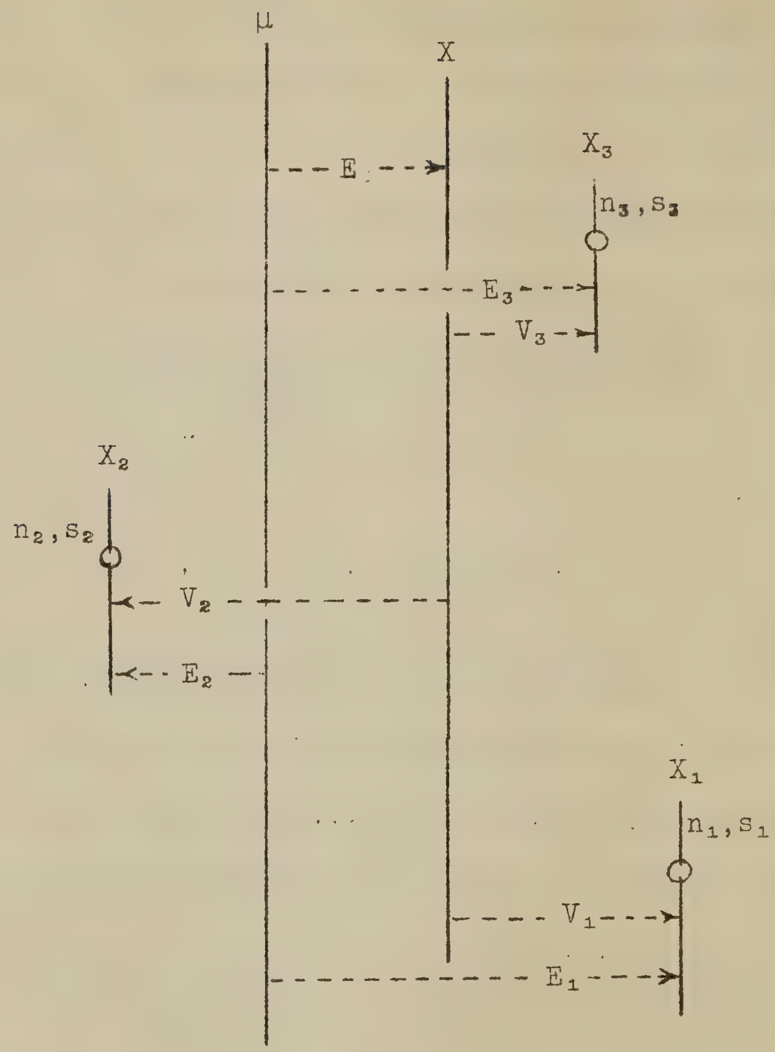

Fig. 2. Three Series of Observations on a Magnitude μ

n_1	observations	have	mean	X_1	and	S. D.	s_1.
n_2	"	"	"	X_2	"	"	s_2.
n_3	"	"	"	X_3	"	"	s_3.

X is the weighted mean of the three series.

The errors and residuals in the individual means X_1, X_2, X_3 are denoted by E_1, E_2, E_3 and V_1, V_2, V_3 respectively. The error in the general mean X is denoted by E. As the figure happens to be drawn, E, E_1, E_3, V_1, and V_3 are positive and E_2 and V_2 are negative, as the arrows indicate. This case of curve fitting is intermediate between the simplest problem shown on page 7 and the more general one on pages 82 and 83.

6. Several samples, all on the same unknown. (a) All observations have the same precision. Let us suppose that n observations of equal weight (equal precision) and all on the same unknown, as for example those of section 3, are arbitrarily subdivided into m samples of $n_1, n_2, \ldots, n_m$ observations. We shall say that

X_1 is the mean of n_1 single observations

X_2 " " " " n_2 " "

$\vdots$ $\vdots$

X_m " " " " n_m " "

Now if single observations have unit weight, then it will follow from Eqs. 14 or 15 that the weights of the m means are

$$w_1 = n_1, \; w_2 = n_2, \; \ldots, \; w_m = n_m$$

We may now consider the m sample means to be m observations of weights respectively $n_1, n_2, \ldots, n_m$, to which the results of section 4 apply. The value of a that minimizes χ^2 is then by Eq. 10,

$$X = \frac{\Sigma wX}{\Sigma w} = \frac{n_1X_1 + n_2X_2 + \ldots + n_mX_m}{n_1 + n_2 + \ldots + n_m} \tag{17}$$

This X is the "weighted mean" of the m samples; it is simply the arithmetic mean of the $n_1 + n_2 + \ldots + n_m$ single observations since they all are of equal weight. Deviations (V_i) reckoned from X make

$$\chi^2 = (1/\sigma^2)\, \Sigma wV^2$$

a minimum, and it is noteworthy that, at the same time, $\Sigma V = 0$.

It should be kept in mind that X is independent of just how the n observations are subdivided, but χ^2 is not. Moreover, both X and χ^2 will vary from one set of observations to another.

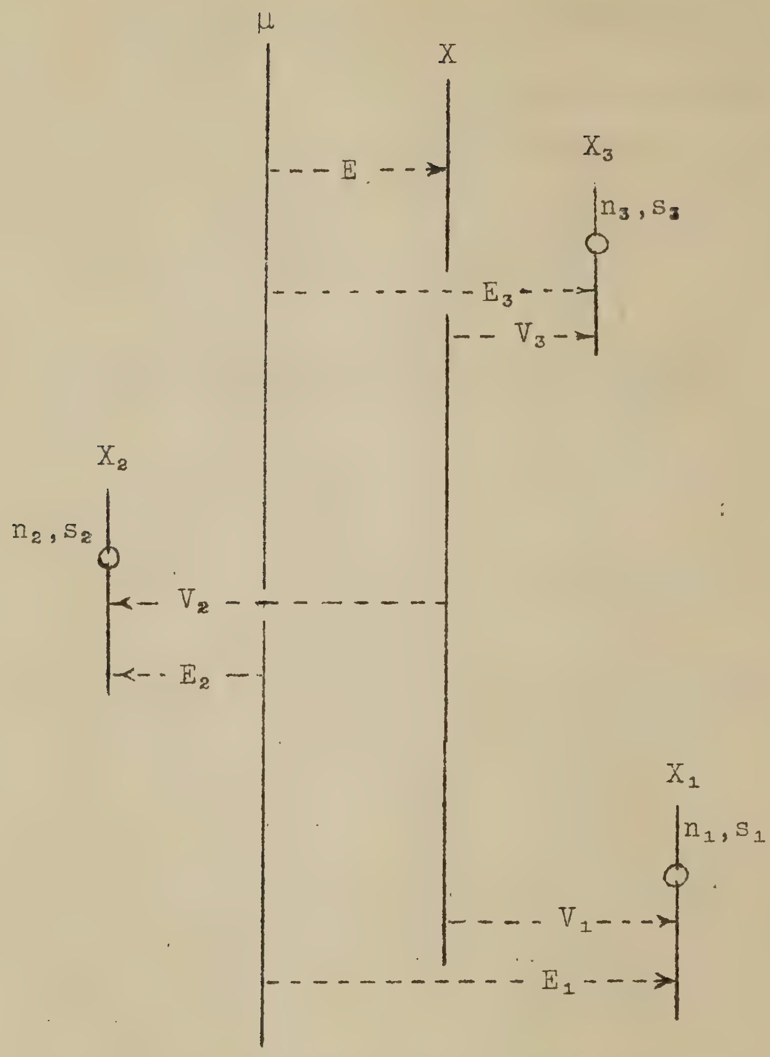

Fig. 2. Three Series of Observations on a Magnitude μ

n_1	observations	have	mean	X_1	and	S. D.	s_1.
n_2	"	"	"	X_2	"	"	s_2.
n_3	"	"	"	X_3	"	"	s_3.

X is the weighted mean of the three series.

The errors and residuals in the individual means X_1, X_2, X_3 are denoted by E_1, E_2, E_3 and V_1, V_2, V_3 respectively. The error in the general mean X is denoted by E. As the figure happens to be drawn, E, E_1, E_3, V_1, and V_3 are positive and E_2 and V_2 are negative, as the arrows indicate. This case of curve fitting is intermediate between the simplest problem shown on page 7 and the more general one on pages 82 and 83.

(b) <u>The precisions of the single observations differ from one sample to another.</u> Suppose that

$$\left.\begin{array}{l} X_1 \text{ is the mean of } n_1 \text{ observations from a population of S.D.} \sigma_1 \text{, the variance of } X_1 \text{ being } \sigma_1^2/n. \\ \vdots \\ X_m \text{ is the mean of } n_m \text{ observations from a population of S.D.} \sigma_m \text{, the variance of } X_m \text{ being } \sigma_m^2/n_m. \end{array}\right\}$$

$\sigma_1, \sigma_2, \ldots, \sigma_m$ need not all be equal. Then we may take

$$\left.\begin{array}{ll} w_1 = \sigma^2/(\sigma_1^2/n_1) = n_1\sigma^2/\sigma_1^2 & \text{for the weight of } X_1 \\ w_2 = \sigma^2/(\sigma_2^2/n_2) = n_2\sigma^2/\sigma_2^2 & \text{" " " " } X_2 \\ \vdots & \\ w_m = \sigma^2/(\sigma_m^2/n_m) = n_m\sigma^2/\sigma_m^2 & \text{" " " " } X_m \end{array}\right\} \tag{18}$$

σ^2 is arbitrary, i.e. the weights are purely relative. Eq. 10 applied to this problem gives

$$X = \frac{\sum wX}{\sum w} = \frac{n_1X_1/\sigma_1^2 + n_2X_2/\sigma_2^2 + \ldots + n_mX_m/\sigma_m^2}{n_1/\sigma_1^2 + n_2/\sigma_2^2 + \ldots + n_m/\sigma_m^2} \tag{19}$$

for the least squares value of a. This X is the weighted mean of the m samples; deviations (V_i) reckoned from it make

$$\chi^2 = (1/\sigma^2)\sum wV^2 \qquad \text{(as defined in Eq. 1)}$$

a minimum (see Fig. 2). At the same time, these residuals give $\sum wV = 0$.

<u>Note</u> that the problem of part (b) reduces to that of part (a) if $\sigma_1 = \sigma_2 = \ldots = \sigma_m$.

<u>Note</u> that σ does not appear in the fraction of Eq. 19, i.e., X is independent of σ. If σ^2 be doubled, all weights would be doubled, but X would be unaltered. Likewise χ^2 in Eq. 20 would be unaltered. (See the exercises in section 5).

Note also that χ^2 can be written

$$\chi^2 = \Sigma \frac{(\text{Discrepancy between } X_i \text{ and } X)^2}{\text{Variance of } X_i \text{ about true mean}} \tag{20}$$

See Ex. 1 of section 5, page 13.

(c) The estimates of σ. On account of the distribution* of χ^2 when the actual sampling (the experimental work) is described by the mathematical model here assumed, namely, normally distributed observations, the mean value of χ^2 in the long run is equal to k, the number of independent residuals or "degrees of freedom".** For the problems of parts (a) and (b), this number is m-1 on account of the fact that there is one relation (Eqs. 17 or 18) between the m residuals and X. The unbiased+ estimate of σ^2 made by external consistency++ is found by

---

* Karl Pearson, Phil. Mag. 50, 157-175, 1900. A paper dealing more specifically with curve fitting of the kind here considered will be found in the J. Amer. Stat. Assoc. 29, 372-382, 1934; see also Phil. Mag. 19, 389-402, 1935.

** The correction for the number of unknowns evaluated (one in this case), and the equivalent of setting the mean value of χ^2 equal to ϕ^2 divided by the number of observed quantities diminished by the number of unknowns evaluated, were set forth by Gauss in his Theoria Combinationis Observationum Erroribus Minimis Obnoxiae, Pars posterior (Göttingen, 1823; vol. 4 of his Werke) Art. 38. This correction is sometimes credited to Bessel, though in my own writings I have cited Encke; but the reference just given, for which I must thank my mentor Dr. G. J. Lidstone, places the originality with Gauss.

+ Unbiased in the sense that its mean value is σ^2; see J. Neyman Washington Lectures (The Graduate School, 1938; $1.25) pp. 131, 132, 135.

++ The terms external and internal consistency were introduced by Birge (Physical Review 40, 207-227, 1932) in a discussion on the mean value of the ratio of these two estimates, the occasion being a paper by Scarborough (Proc. Nat. Acad. Sci. 15, 665-668, 1929) denying the validity of the external estimate. The comparison of the two estimates (p. 21) is an application of the "analysis of variance", the essential features of which have long been recognized by physical scientists; see for example A. de Forest Palmer Theory of Measurements (McGraw-Hill, 1912) pp. 66-71.

arbitrarily saying that χ^2 has its mean value k, in other words that σ satisfies Eq. 1a, page 5, whence comes the estimate

$$\sigma^2(\text{ext}) = \phi^2/k \tag{21}$$

From this we get, for the problem of section 6b,

$$\sigma^2(\text{ext}) = (\Sigma wV^2)/k = \frac{\Sigma(n_i\sigma^2/\sigma_i^2)(X_i - X)^2}{m - 1} \tag{22}$$

This estimate is made from the external consistency of the data, i.e., from the fit of the "curve" X = a. What we do in making the estimate σ(ext) is to say arbitrarily that χ^2 does equal k. This is equivalent to saying that P(χ) is about $\frac{1}{2}$--not exactly $\frac{1}{2}$ because of the skewness of the χ^2 distribution, which, however, gradually disappears with increasing k.

If we are not positive that all m samples came from populations having coincident means, we should have as an alternate hypothesis that the m population means $\mu_1, \mu_2, \ldots, \mu_m$ are not all identical. Now if one or more of them really are not equal to the others, σ^2(ext) is raised, on the average, to some value higher than σ^2; consequently in examining the hypothesis that $\mu_1 = \mu_2 = \ldots = \mu_m$, we should be interested in knowing if σ^2(ext) is significantly greater than σ^2, or, what is the same thing, if χ^2 is significantly higher than k. This can be ascertained by looking up P(χ) in tables of chi-square. Of course, χ^2 can not be computed nor compared with k unless σ is known. Or, to use Fisher's tables of z, one would set

$$z = \tfrac{1}{2} \ln [\sigma^2(\text{ext})/\sigma^2] \tag{23}$$

and look up P(z) with Fisher's n_1 as m - 1, and with n_2 equal to infinity, since σ^2 is here assumed known.

Now the

$$\text{Wt. of } X \equiv w_X = \Sigma n_i \sigma^2/\sigma_i^2 \tag{24}$$

whence the

$$(\text{Est'd. S.E. of } X)^2{}_{\text{ext}} = \sigma^2(\text{ext})/w_X$$

$$= \Sigma wV^2/(m-1)w_X = \phi^2/(m-1)w_X \tag{25}$$

wherein w_X is to have the value given in Eq. 24.

There is also the estimate of σ made from the <u>internal consistency</u>* of the data, i.e. from the consistency of the observations within samples.** This is+

$$\sigma^2(\text{int}) = \frac{n_1 s_1^2 + n_2 s_2^2 + \ldots + n_m s_m^2}{n_1 + n_2 + \ldots + n_m - m} \tag{26}$$

whence the

$$(\text{Est'd S.E. of } X)^2{}_{\text{int}} = \sigma^2(\text{int})/w_X \tag{27}$$

wherein w_X has the value given in Eq. 24.

---

* See the reference to Birge on page 18.

** It is assumed that there is more than one observation in each sample, whereupon the estimate by internal consistency is actually possible. Unfortunately as data are too often taken, there may be but one observation at each point, and the estimate of σ by internal consistency is then not a possibility. This is to be regarded as a technical fault of experimentation, a matter of design.

\+ See, for example, Eq. 67 in Deming and Birge's <u>Statistical theory of Errors</u> (obtainable from The Graduate School, 35 cents) p. 158.

(d) Comparison of the two estimates. As was mentioned in part (c), the estimate σ(ext) is valid only if the m populations have coincident means; if any two of the means $\mu_1, \mu_2, \ldots, \mu_m$ are unequal, σ^2(ext) is, on the average, raised above σ^2. But, in contrast, the estimate σ(int) is unaffected by inequalities among the means of the populations; so long as σ remains the constant S.D. of all of them, the average value of σ^2(int) is still σ^2. It follows that a statistical test of the hypothesis $\mu_1 = \mu_2 = \ldots = \mu_m$ is to examine the ratio of the two estimates. To do this, we may follow Fisher and take

$$z = \tfrac{1}{2} \ln [\sigma^2(\text{ext})/\sigma^2(\text{int})] \tag{28}$$

and look in Fisher's tables to see if z is significantly different from 0. (In doing this, we use m-1 for Fisher's n_1, and $n_1 + n_2 + \ldots + n_m - m$ for his n_2). If z is found to lie beyond the 1 percent limit, we say there is "statistical evidence" that the data are not homogeneous, or that not all the μ_i are equal; in other words, that the curve

$$X_i = a \tag{29}$$

is not a good fit. Such a calculation takes account only of the numerical data, and any conclusions therefrom must of course not be taken too seriously, but rather should be weighed along with the considerations of the experimental work, and previous experience.

The student will realize that a comparison of two estimates of the same σ is an "analysis of variance" (cf. footnote on page 18). R. A. Fisher's Statistical Methods for Research Workers (Oliver & Boyd) contains many examples, a particularly good one being Example 42, "Test of straightness of regression line".

It must be kept in mind that a new set of data will give new values of σ(ext) and σ(int), hence new values of $P(\chi)$ and $P(z)$. It is these variations that statistical tests lay odds on, the assumption being that the experiment is "under control"; see Lecture II of Neyman's Washington Lectures (The Graduate School, 1938); also Shewhart's lectures (The Graduate School, 1938).

7. Another simple problem--the slope of a line that is known to pass through the origin. In general from section 1 we may say that

$$\chi^2 = \frac{1}{\sigma^2}\Sigma(w_x V_x^2 + w_y V_y^2)$$

So far, we have seen only special cases of this--special in the sense that the error was all in the x coordinates of the points. Now we take

(a) The y coordinates subject to error; x free of error. If w_i denotes the weight of y_i, then

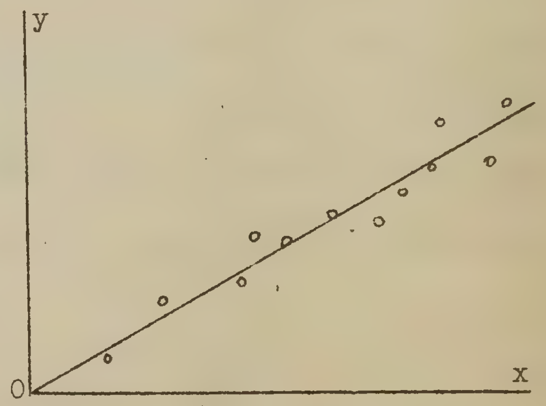

Fig. 3. A line known to pass through the origin, the slope to be estimated from the observed points.

$$\phi^2 = \Sigma w_i(y_i - bx_i)^2 \tag{30}$$

is to be minimized. y_i is observed, and bx_i is its calculated value; the difference between them is the vertical or y residual at the i-th point. We differentiate ϕ^2 and obtain

$$d\phi^2/db = -2\Sigma w_i x_i(y_i - bx_i)$$

Set equal to zero, this gives

$$b\Sigma w_i x_i^2 = \Sigma w_i x_i y_i \tag{31}$$

whence

$$b = \Sigma wxy/\Sigma wx^2 \tag{32}$$

Note that here, Σ w res $\neq$ 0, but that Σ w.x.res = 0. How does this come about? (Cf. remark 3, p. 126 for further reference).

Special case: w_i proportional to $1/x_i$; i.e. the S.E.^{2} of y_i is proportional to x_i. Then Eq. 32 gives

$$b = \Sigma\, y/\Sigma x \tag{33}$$

Note that here, Σ w.res = 0.

Another special case: All weights equal, i.e. all y_i have the same S.E. Then Eq. 32 gives

$$b = \Sigma xy/\Sigma x^2 \tag{34}$$

This is perhaps a more usual one than the previous special case.

Note carefully the distinction between Eqs. 33 and 34. See how in Eqs. 32 and 34 a point has more influence if it is far outlying, the influence being proportional to x. Is this reasonable? But in Eq. 33 all points are equal in influence, near or far. Why should this be? or rather, under what conditions would Eq. 33 be a reasonable result?

(b) The tabular calculation of b and its weight. This will be similar to the tabulation in section 4b (q. v.).

Suppose we write

No.	b	=	1	C	
I	Σwx^2		Σwxy	1	
2	...		Σwy^2	0	
3	...		$-\frac{(\Sigma wxy)^2}{\Sigma wx^2}$	...	(Gotten by multiplying I through by $-\Sigma wxy/\Sigma wx^2$)
II	...		$\Sigma wy^2 - \frac{(\Sigma wxy)^2}{\Sigma wx^2}$	...	(By adding 2 and 3)

No. I solved with the "1" column gives b as in Eq. 32.

" I " " " C " " $b = 1/\Sigma wx^2$, which means that

$$w_b = \Sigma wx^2 \qquad (35)$$

No. II gives the minimized value of ϕ^2 or of $\Sigma w_i(y_i - bx_i)^2$ directly (found in the "1" column), without the intermediate step of calculating b and the individual residuals and their squares. The student should follow this carefully, especially after studying section 17ff.

Here we have

$$\sigma^2(\text{ext}) = \phi^2 / (m-1) \qquad (36)$$

and the

$$(\text{Est'd. S.E. of b})^2_{\text{ext}} = \frac{\phi^2}{(m-1)w_b} = \frac{\phi^2}{(m-1)\Sigma wx^2} \qquad (37)$$

In order to apply the t test to see if there is "statistical evidence" that the calculated value of b is significantly different from some theoretical value, say B, we should write

$$t = \frac{|B-b|}{\text{Est'd S.E. of b}} \qquad (38)$$

and make the t test with Fisher's n equal to our m-1. The region of rejection in the t distribution is to be chosen with due regard to admissible alternative slopes, which may be greater or less than B. In the denominator of Eq. 38 we may use the estimate made by external consistency, or that made by internal consistency (section 6c). If σ(int) were used in place of σ(ext) in Eq. 37, then we should have

$$(\text{Est'd S.E. of } b)^2_{\text{int}} = \sigma^2(\text{int})/w_b = \sigma^2(\text{int})/\Sigma wx^2 \qquad (39)$$

This would replace the denominator of Eq. 38, and the number of degrees of freedom (Fisher's n) would be the total number of observations diminished by the number m.

The testing of parameters is closely tied up with the theory of confidence intervals for any function of the parameters; see page 116.

(c) Now let the x coordinates be subject to error, y free of error.

w_i will now denote the weight of x_i. In place of Eq. 30 we now have

$$\phi^2 = \Sigma w_i (x_i - y_i/b)^2 \qquad (40)$$

since here the y residuals are zero, and ϕ^2 is made up by squaring the x residuals. By differentiation

$$d\phi^2/db = +(2/b^2)\Sigma w_i y_i (x_i - y_i/b) \qquad (41)$$

Set equal to zero this gives

$$b \, \Sigma\, wxy = \Sigma\, wy^2, \text{ or } b = \Sigma\, wy^2/\Sigma\, wxy \qquad (42)$$

Note the distinction between Eq. 32 and 42.

Exercise 1. If b in Eq. 42 be distinguished as b', prove that between Eqs. 32 and 42 there exists the relation

$$b/b' = r^2 \tag{43}$$

r being the correlation between x and y. Hence $b = b'$ only if the points all lie exactly on a line, since then $r^2 = 1$ and $b = b'$.

Exercise 2. Prove by making up a tabulation for b' similar to the one on page 24 for b, or by any other method, that the weight of b' (the b in Eq. 42) is $\sum wxy$; in symbols

$$w_{b'} = \sum wxy \tag{44}$$

Exercise 3. Prove by the same tabulation or otherwise that in section 7c the minimized ϕ^2 is

$$\phi^2 = \sum wx^2 - (\sum wxy)^2/\sum wy^2 \tag{45}$$

8. The propagation of error. Let F be a function of x, y, z. Then if x, y, z are in error by Δx, Δy, Δz, F will be in error by the amount ΔF, where by Taylor's series

$$\Delta F = F_x\Delta x + F_y\Delta y + F_z\Delta z + \ldots \tag{46}$$

wherein

$$F_x = \frac{dF}{dx},\quad F_y = \frac{dF}{dy},\quad F_z = \frac{dF}{dz}$$

Formula 46 is often called the propagation of error.

Now square each side of Eq. 46 and get

$$\Delta F^2 = (F_x\Delta x)^2 + (F_y\Delta y)^2 + (F_z\Delta z)^2 + 2F_xF_y\Delta x\Delta y$$
$$+ 2F_xF_z\Delta x\Delta z + 2F_yF_z\Delta y\Delta z + \ldots$$

Imagine F_x, F_y, F_z to be nearly constant as Δx, Δy, and Δz take on all possible values.* Then let each term be replaced by its average value; the result is

$$\sigma_F^2 = (F_x\sigma_x)^2 + (F_y\sigma_y)^2 + (F_z\sigma_z)^2 + 2(F_xF_y\sigma_x\sigma_y r_{xy}$$
$$+ F_xF_z\sigma_x\sigma_z r_{xz} + F_yF_z\sigma_y\sigma_z r_{yz}) \tag{47}$$

where σ_x^2 = variance of x, r_{xy} the correlation between x and y, etc This formula (also the simplified form in Eq. 48 when it applies) is called the propagation of mean square error, or the propagation of variance.

The terms in parenthesis are zero if the errors in x, y, and z are independent, i. e. uncorrelated. In such a situation Eq. 47 reduces to

$$\frac{1}{w_F} = \frac{F_xF_x}{w_x} + \frac{F_yF_y}{w_y} + \frac{F_zF_z}{w_z} \tag{48}$$

---

* In practice it may safely be assumed that the ranges of variation in Δx, Δy, and Δz are not large, wherefore the constancy of F_x, F_y, and F_z is usually not a difficulty. It is moreover presumed that σ_x^2, σ_y^2, and σ_z^2 actually do exist (as would not be the case, for instance, in a Cauchy distribution of the errors in x, y, and z).

It is interesting to see that if F be taken as the mean ($\bar{x}$) of the n independent observations $x_1, x_2, \ldots, x_n$ each of S. E. σ, then Eq. 48 leads to the well known expression

$$\sigma_{\bar{x}}^2 = \sigma^2/n$$

as was taken for granted on page 12. This, however, does not tell us that if the individual observations are normally distributed, the mean $\bar{x}$ is also--this fact must be obtained otherwise. Eqs. 47 and 48 are in fact independent of any assumption concerning the distributions of the errors in x, y, z, and F, provided the S. Es. σ_x^2 etc. actually exist, as was stipulated in the footnote on the preceding page.

Exercise 1. The mean square error of the sum or difference of two numbers having equal precisions is twice the mean square error of either alone (assumed independent). The r.m.s. error (S.E.) of the sum or difference is $\sqrt{2}$ times that of either alone.

Exercise 2. If u is a linear function of the independent variables x, y, and z, say

$$u = ax + by + cz$$

then the r.m.s. errors are related by the equation

$$\sigma_u^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + c^2\sigma_z^2$$

A special case is exercise 1.

Exercise 3. If $u = axyz$, then

$$\sigma_u^2 = u^2\{(\sigma_x/x)^2 + (\sigma_y/y)^2 + (\sigma_z/z)^2\}$$

or
$$(\sigma_u/u)^2 = (\sigma_x/x)^2 + (\sigma_y/y)^2 + (\sigma_z/z)^2$$

which interpreted says that the squares of the percentage errors are additive.

Exercise 4. If $u = axy/vz$, then

$$(\sigma_u/u)^2 = (\sigma_x/x)^2 + (\sigma_y/y)^2 + (\sigma_v/v)^2 + (\sigma_z/z)^2$$

the squares of the percentage errors being again additive.

Exercise 5. If $u = ax^\alpha y^\beta z^\gamma$, then

$$(\sigma_u/u)^2 = (\alpha\sigma_x/x)^2 + (\beta\sigma_y/y)^2 + (\gamma\sigma_z/z)^2$$

Here the percentage errors are increased by the factors α, β, γ. (Exercises 3 and 4 are special cases of this).

Exercise 6. If $A = \pi r^2$ (A the area and r the radius of a circle), then if an error Δr be committed in measuring r, the corresponding error in the area is closely

$$\Delta A = 2\pi r\ \Delta r = 2(A/r)\Delta r$$

$$\frac{\Delta A}{A} = 2\ \frac{\Delta r}{r}$$

An error of 1 percent in the radius means about 2 percent error in the area. Also

$$\sigma_A = 2\ \sigma_r \quad \text{and} \quad w_A = (1/4)w_r$$

(special case of Ex. 5).

Exercise 7. For the conditions of exercises 4 and 5 the relations between the weights are respectively

$$1/u^2 w_u = 1/x^2 w_x + 1/y^2 w_y + 1/z^2 w_z$$

and

$$1/u^2 w_u = \alpha^2/x^2 w_x + \beta^2/y^2 w_y + \gamma^2/z^2 w_z$$

Exercise 8. (a) Let* $Y = \ln y$, then

$$\sigma_Y^2 = (1/y^2)\sigma_y^2$$

$$w_Y = y^2 \, w_y$$

(This result is very important; see exercise 18 in section 19)

(b) If y is in error by the amount δy, log y is in error $(.4343\ \delta y)/y$.

(c) If $Y = \log y$, then

$$\sigma_Y^2 = (0.4343\ \sigma_y/y)^2$$

$$w_Y = 5.3\ y^2\ w_y$$

Exercise 9. Let $u = ae^{bx}$, then

$$\sigma_u^2 = b^2 u^2 \sigma_x^2$$

Exercise 10. The period of a simple pendulum is $T = 2\pi\sqrt{(L/g)}$. Show that if the length L is too long by one-tenth of a percent, the clock will lose about 44 seconds per day.

Exercise 11. (a) Prove that if F is a function of x, and x a function of t, that

$$F_x F_x / w_x = F_t F_t / w_t$$

where F_x denotes dF/dx, and F_t denotes dF/dt.

(b) Prove that

$$1/w_x = (dx/dt)^2 / w_t$$

(These results are useful in section 19).

---

* The abbreviation ln is used here for "logarithme naturel," as is common in Europe and among chemists everywhere. The abbreviation log will be used for logarithms to base 10.

II. MORE COMPLICATED PROBLEMS

9. The general problem in least squares*. The fundamental notions involved are the true curve (unknown), the observed points, and the calculated curve. We shall deal with the following quantities.

Observed coordinates**:	$X_1, X_2, \ldots, X_n;\ X_{n+1}, X_{n+2}, \ldots, X_{2n}$
Weights:	$w_1, w_2, \ldots, w_n;\ w_{n+1}, w_{n+2}, \ldots, w_{2n}$
True values (unknown):	$\xi_1, \xi_2, \ldots, \xi_n;\ \xi_{n+1}, \xi_{n+2}, \ldots, \xi_{2n}$
Calculated coordinates:	$x_1, x_2, \ldots, x_n;\ x_{n+1}, x_{n+2}, \ldots, x_{2n}$
Residuals (Obs.-Calc.):	$V_1, V_2, \ldots, V_n;\ V_{n+1}, V_{n+2}, \ldots, V_{2n}$

In dealing with curve fitting, X_{n+1} will take the place of Y_1, X_{n+2} the place of Y_2, etc., for convenience in writing summations (see Eq. 49 for instance). If the problem is not one of curve fitting, as for example if it were one involving geometrical conditions with no adjustable parameters, we need not think of the observations as being coordinates of points, and in such problems the quantities listed above (X, w, ξ, x) take subscripts from 1 to n and no further, the subscripts n+1, n+2, ... , 2n not then being used.

---

* The development from here on is an amplification of three papers that appeared in the London Philosophical Magazine. The references are vol. 11, pp. 146-158, 1931; vol. 17, pp. 804-829, 1934; vol. 19, pp. 389-402, 1935.

** The use of the capital letters here is consistent with section 6, since the numbers listed as observed coordinates may actually be the means of several observations. This is taken care of in the weights.

In any case, the principle of least squares requires that*

$$\phi^2 = \Sigma w_i V_i^2 = \Sigma w_i (X_i - x_i)^2 \tag{49}$$

shall be made a minimum with respect to x_1, x_2, ... But this is not a simple problem in maximum and minimum values; the adjusted values x_i are related to one another, e.g. in the case of measurements on the three angles of a plane triangle, we should require that $x_1 + x_2 + x_3 = 180°$ (see section 11a). What is more, in curve fitting, the conditions on the adjusted values x_i involve unknown parameters α, β, γ whose estimates a, b, c are sought. In the problem of section 3, for instance, the adjusted values of the x coordinates of the n points were all required to be equal to a, which was then evaluated as $\bar{x}$ (Fig. 1) to make the sum of the squares of the residuals a minimum. So to take care of the general case we shall suppose that the adjusted values x_i are subject to ν conditions, designated by

$$\left.\begin{array}{ll} F^1\,(x_1, x_2, \ldots, x_{2n}; a, b, c) & = 0 \\ F^2\,(\quad " &) = 0 \\ \vdots & \vdots \\ F^{\nu'}(\quad " &) = 0 \end{array}\right\} \begin{array}{l} \nu \text{ equations} \\ \quad \text{for} \\ \nu \text{ conditions} \end{array} \tag{50}$$

Different sorts of problems are distinguished by the different kinds of conditions that the adjusted quantities x_i, and a, b, c are subjected to. They are all conveniently handled as one problem, because there is only one principle of least squares--the minimizing of χ^2.

---

* The sign Σ will denote summation with i running over all observations.

These will be referred to as the <u>conditions</u>--the conditions imposed upon the adjusted values x_i and the estimated parameters a, b, c (if any). The assumption behind the whole process is that the conditions would all be satisfied exactly by the true (unknown) quantities being measured, and the true parameters α, β, γ.

Let derivatives of F^h be denoted by subscripts* as in Eqs. 46-48; e. g. let

$$\begin{aligned} F_1^h &= dF^h/dx_1, \quad F_2^h = dF^h/dx_2, \text{ etc.} \\ F_a^h &= dF^h/da, \quad F_b^h = dF^h/db, \text{ etc.} \end{aligned} \tag{51}$$

wherein, after differentiation, the observations are to be inserted, along with the best available approximate values a_o, b_o; c_o (infra). In other words, F_1^h is to be a number representing our best guess ** at the value of the derivative dF^h/dx_1, and similarly for the other derivatives in Eqs. 51.

We also write+

$$F_o^h = F^h\ (X_1, X_2, \ldots , X_{2n}; a_o, b_o, c_o) \tag{52}$$

---

* Denoting differentiations by subscripts is a common practice in mathematics. The superscript h, which may be 1, or 2, ... , or v, merely distinguishes one F (i.e. one condition) from another.

** If our best guess is too far off, a second adjustment may be required, but this rarely happens in practice. See the quotation from Gauss in Ex. 4 of section 19, pp. 124-125.

\+ Mr. Jesse H. Buell of the Forest Service has kindly directed my attention to an inconsistency in notation; the subscripts 1, 2, 3, ... , a, b, c in Eqs. 51 denote differentiations, but the subscript o in Eq. 52 does not. I fear it will have to stand. Fortunately, confusion does not ordinarily arise, especially after the inconsistency is recognized.

F_o^1 is a small number, the amount by which the condition $F^1 = 0$ fails to be satisfied by the observed values X_1, X_2, ... and the approximations a_o, b_o, c_o. Similar statements hold for F_o^2, F_o^3, ... , F_o^v.

a_o, b_o, c_o are approximate values of a, b, c. They can usually be arrived at somehow, as by putting three of the F_o^h equal to zero (method of selected points*). The final adjusted values a, b, c will be independent of the approximations a_o, b_o, c_o, except that occasionally these approximations must not be too rough; see section 13. Each F_o^h would be zero except for errors of observation and the consequent impossibility of choosing a_o, b_o, c_o to satisfy all the conditions exactly. By definition, $v_a = a_o - a$, $v_b = b_o - b$, $v_c = c_o - c$.

Now let the conditions be made linear in the residuals V_1, V_2, ... , V_{2n}, v_a, v_b, v_c, by expanding Eqs. 50, retaining only the first powers of the residuals**, and remembering that $x_i = X_i - V_i$, $a = a_o - v_a$, etc.

---

* The "method of selected points" could be considered a method of curve fitting were it not about the worst conceivable method, and hardly worthy of the name. It throws away all the information contained in the points not selected; yet this much can be said, if there were no errors in any of the points, it would yield correct results. Generally speaking, it is well to select the points as far apart as possible when adopting the plan. There are of course other methods of finding approximate values: there are graphical methods, the method of averages (Norman Campbell's method of zero-sum, Phil. Mag. 39, 177-194, 1920), and Cauchy's method Comptes rendus 25, 650, 1847. Frequently one will have good enough approximations from previous experience. When I speak of approximations by any of these methods, I do not imply that they are inferior to least squares--they are only different. The values a_o, b_o, c_o that any one of them gives may be very close to the least squares values a, b, c, and thus be useful in the least squares solution, when such seems desirable.

** The problem of a straight line with no error at all in one of the coordinates (Exercises 1 and 7 in section 19) is one in which there are no squares and higher powers of the residuals to neglect, hence no discrepancies of the kind mentioned. The simple example of the triangle in section 11 will be another. On rare occasions the residuals

The result is

$$\Sigma F_i^h V_i + F_a^h v_a + F_b^h v_b + F_c^h v_c = F_o^h \qquad \begin{matrix} h = 1,2,\ldots,\nu \\ \nu \text{ equations} \end{matrix} \qquad (53)$$

These are called the <u>reduced conditions</u>. They are equivalent to Eqs. 50, except for small discrepancies arising from the neglect of higher powers of the residuals in the expansion.

Now if ϕ^2 is at its minimum value, and if any or all of the residuals then undergo small variations (expressed by δ), the variation in ϕ^2 will be zero to within higher powers of the variations in the residuals; in other words

$$\frac{1}{2} \cdot \delta\phi^2 = \Sigma w_i V_i \; \delta V_i = 0, \qquad \text{one equation} \qquad (54)$$

The variations δV_1, δV_2, etc. are not arbitrary but must always permit V_1, V_2, etc. to satisfy the condition Eqs. 50 or their equivalent (53). So by differentiating Eqs. 53 we find that

$$\Sigma F_i^h \; \delta V_i + F_a^h \; \delta v_a + F_b^h \; \delta v_b + F_c^h \; \delta v_c = 0 \qquad \begin{matrix} h = 1, 2, \ldots,\nu \\ \nu \text{ equations} \end{matrix} \qquad (55)$$

Now multiply Eq. 55 thru by $-\lambda_h$, an arbitrary multiplier;

$$-\lambda_h(\Sigma F_i^h \; \delta V_i) - \lambda_h F_a^h \; \delta v_a - \lambda_h F_b^h \; \delta v_b - \lambda_h F_c^h \; \delta v_c = 0, \qquad \begin{matrix} h = 1,2,\ldots, \nu \\ \nu \text{ equations} \end{matrix} \qquad (56)$$

Add Eqs. 54 and 56 and collect coefficients of the variations δ:

$$\Sigma(w_i V_i - [\lambda_h F_i^h]) \, \delta V_i - [\lambda_h F_a^h] \, \delta v_a - [\lambda_h F_b^h] \, \delta v_b$$
$$- [\lambda_h F_c^h] \, \delta v_c = 0 \qquad \text{one equation} \qquad (57)$$

---

may be so large that the neglected terms invalidate the reduced conditions (53), in which event, in general, no systematic solution is available. An exception is the straight line under certain circumstances of weighting; see Exercise 6 of section 19.

In Eq. 57 there are two kinds of summations--there is the summation Σ in which i runs over all observations; and there is also the summation over h, in which h runs from 1 to ν, i.e. over all conditions. The latter summation will be denoted by Gauss' [ ].

This equation contains 2n+3 variations*, δV_1, δV_2, δV_3, ... , δV_{2n}, δv_a, δv_b, δv_c. But only 2n+3-ν of them are arbitrary on account of Eqs. 55, which are ν in number. Let λ_1, λ_2, ... , λ_ν be so chosen that ν of the coefficients in Eq. 57 vanish; then the coefficients of the variations in the remaining 2n+3-ν terms must also vanish, because they are used with an equal number of variations, each of which is arbitrary**. Then all the coefficients in Eq. 57 vanish, that is,

$$V_i = \frac{1}{w_i}[\lambda_h F_i^h] \qquad \text{2n equations, } i = 1, 2, \ldots , 2n \tag{58}$$

$$[\lambda_h F_a^h] = 0, \quad [\lambda_h F_b^h] = 0, \quad [\lambda_h F_c^h] = 0 \tag{59 a,b,c}$$

The ν Lagrange multipliers (λ) are no longer arbitrary, having been chosen so as to cause ν of the coefficients in Eq. 57 to vanish

---

* Here the number of parameters is taken as 3. If there were p parameters, there would be 2n+p variations. For practice, the student should write out Eqs. 54-61 with (e.g.) n=3 and ν=2 with (say) 2 parameters. There is no other way to gain familiarity with the development.

** This is the method of Lagrange multipliers; see his Mecanique Analytique (1811) tome 1, p. 74; or Benjamin Williamson Differential Calculus (Longmans, 1893) Ch. 11. The least squares problem without parameters was worked out by Gauss. He called his multipliers correlata, not mentioning Lagrange. Many texts in least squares use the term "correlates" or "correlatives" in this connection, but none that I have seen makes any mention of Lagrange. The reference to Gauss is his Supplementum Theoriae Combinationis Observationum Erroribus Minimus Obnoxiae (Göttingen, 1826; Werke, vol. 4) art. 11.

(vide supra). Their values are given by Eqs. 61.

Now substitute $\frac{1}{w_i}[\lambda_h F_i^h]$ for V_i in the reduced conditions (53). Collect the coefficients of $\lambda_1, \lambda_2, \ldots, \lambda_v, v_a, v_b, v_c$, and in so doing set

$$L_{rs} = \frac{F_1^r F_1^s}{w_1} + \frac{F_2^r F_2^s}{w_2} + \ldots + \frac{F_{2n}^r F_{2n}^s}{w_{2n}} = L_{sr} \qquad (60)$$

The following system of equations results. I call them the "general normal equations". For convenience only the coefficients are tabled, the unknown being written across the top. On the left of the equality sign, each coefficient is to be multiplied by the unknown appearing above it, the plus sign between terms being understood. On the right, each F_o is multiplied by unity, hence the heading "1" for that column.

The general normal equations

λ_1	λ_2	λ_3	...	λ_v	v_a	v_b	v_c	= 1	
L_{11}	L_{21}	L_{31}	...	L_{v1}	F_a^1	F_b^1	F_c^1	F_o^1	
L_{12}	L_{22}	L_{32}	...	L_{v2}	F_a^2	F_b^2	F_c^2	F_o^2	
L_{13}	L_{23}	L_{33}	...	L_{v3}	F_a^3	F_b^3	F_c^3	F_o^3	(61)
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
L_{1v}	L_{2v}	L_{3v}	...	L_{vv}	F_a^v	F_b^v	F_c^v	F_o^v	
F_a^1	F_a^2	F_a^3	...	F_a^v	0	0	0	0	(59a)
F_b^1	F_b^2	F_b^3	...	F_b^v	0	0	0	0	(59b)
F_c^1	F_c^2	F_c^3	...	F_c^v	0	0	0	0	(59c)

Note that since $L_{rs} = L_{sr}$, the coefficients below the unknowns are symmetrical about the main diagonal.

The general normal equations can be solved for the $\nu + 3$ unknowns written across the top: special methods of solution will be taken up in sections 12b, 17, and 20, but for the present we shall only note that once the residuals v_a, v_b, and v_c are found, the final (adjusted) values of the parameters are obtained by subtracting the residuals from the approximate values, thus:

$$a = a_0 - v_a, \quad b = b_0 - v_b, \quad c = c_0 - v_c \tag{62}$$

It may not be superfluous to remind the reader that the final adjusted values of a, b, and c are not dependent on the approximations a_0, b_0, and c_0, provided these approximations are within reason. To be specific, if a_0, b_0, and c_0 are changed slightly, and a second solution made, the residuals v_a, v_b, and v_c will turn out to be slightly different--just enough in fact to bring a, b, and c into agreement with the first solution. Sometimes the approximations need not be close, yet the solution is still valid; examples will be found in section 19, for instance on pages 119 and 133, and elsewhere.

The solution of the general normal equations yields also numerical values for the Lagrange multipliers $\lambda_1, \lambda_2, \ldots, \lambda_\nu$, which through Eq. 58 enable the residuals V_i to be calculated. The observations X_i are then adjusted by subtracting the residuals, thus,

$$x_i = X_i - V_i \qquad \text{(calc'd = obs'd - residual)} \tag{63}$$

These adjusted quantities x_i, along with the adjusted parameters found by Eqs. 62, will satisfy the v conditions expressed by Eqs. 50, page 32, or their equivalent, Eqs. 53, page 35.

Note by Eq. 58 that the residual V_i is inversely proportional to the weight w_i of the corresponding observation. Does this seem reasonable? If any observation is relatively infallible, having $w = \infty$, then its residual $V_i = 0$; i. e., there is no correction. In curve fitting, for example, it sometimes happens that all the x coordinates are free of error; the corresponding residuals are then 0, and the calculated values of x are the same as the observed.

10. *Short expressions for ϕ^2. The normal equations are really normal, their coefficients symmetrical and positive definite.*

By definition, $\phi^2 = \sum w_i V_i^2$ (49)

Then by making use of Eq. 58,

$$\phi^2 = \sum \frac{1}{w_i}[\lambda_h F_i^h]^2 = \frac{1}{w_1}(\lambda_1 F_1^1 + \lambda_2 F_1^2 + \dots + \lambda_\nu F_1^\nu)^2 +$$

$$\frac{1}{w_2}(\lambda_1 F_2^1 + \lambda_2 F_2^2 + \dots + \lambda_\nu F_2^\nu)^2 + \frac{1}{w_3}(\lambda_1 F_3^1 + \lambda_2 F_3^\nu + \dots$$

$$+ \lambda_\nu F_3^\nu)^2 + \dots + \frac{1}{w_{2n}}(\lambda_1 F_{2n}^1 + \lambda_2 F_{2n}^2 + \dots + \lambda_\nu F_{2n}^\nu)^2$$

$$= L_{11}\lambda_1^2 + L_{22}\lambda_2^2 + \dots + L_{\nu\nu}\lambda_\nu^2 + 2(L_{12}\lambda_1\lambda_2 + \dots)$$

Another way of writing this is the following:

$$\begin{array}{c|cccccccc}
 & \lambda_1 & \lambda_2 & \lambda_3 & \dots & \lambda_\nu & v_a & v_b & v_c \\
\hline
\lambda_1 & L_{11} & L_{21} & L_{31} & \dots & L_{\nu 1} & F_a^1 & F_b^1 & F_c^1 \\
\lambda_2 & L_{12} & L_{22} & L_{32} & \dots & L_{\nu 2} & F_a^2 & F_b^2 & F_c^2 \\
\lambda_3 & L_{13} & L_{23} & L_{33} & \dots & L_{\nu 3} & F_a^3 & F_b^3 & F_c^3 \\
\vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots \\
\lambda_\nu & L_{1\nu} & L_{2\nu} & L_{3\nu} & \dots & L_{\nu\nu} & F_a^\nu & F_b^\nu & F_c^\nu \\
v_a & F_a^1 & F_a^2 & F_a^3 & \dots & F_a^\nu & 0 & 0 & 0 \\
v_b & F_b^1 & F_b^2 & F_b^3 & \dots & F_b^\nu & 0 & 0 & 0 \\
v_c & F_c^1 & F_c^2 & F_c^3 & \dots & F_c^\nu & 0 & 0 & 0
\end{array} \qquad (64)$$

$= \lambda_1 F_o^1 + \lambda_2 F_o^2 + \lambda_3 F_o^3 + \dots + \lambda_\nu F_o^\nu = [\lambda_h F_o^h]$ by Eqs. 61, page 37.

I. e. $\phi^2 = [\lambda_h F_o^h]$ (65)

For special expressions of ϕ^2 in curve fitting, see pages 109-111.

Since $\phi^2 > 0$, the quadratic form (64) is positive definite;* no matter what values be given to $\lambda_1, \lambda_2, \ldots, \lambda_\nu, v_a, v_b, v_c$, the quadratic form (64) can not be negative. The symmetry of the general normal equations (section 9) has already been noted; hence they are really <u>normal</u>.

11. Geometrical conditions, or no parameters. (a) <u>The normal equations</u>. If there are no adjustable parameters, then $F_a^h, F_b^h, F_c^h, v_a, v_b, v_c$, do not exist, and of the general solution (section 9) there is left only the coefficients in the upper left-hand corner:

$$\begin{array}{cccccc}
\lambda_1 & \lambda_2 & \lambda_3 & \cdots & \lambda_\nu & = 1 \\
\hline
L_{11} & L_{12} & L_{13} & \cdots & L_{1\nu} & F_o^1 \\
 & L_{22} & L_{23} & \cdots & L_{2\nu} & F_o^2 \\
 & & L_{33} & \cdots & L_{3\nu} & F_o^3 \\
 & & & & \vdots & \vdots \\
 & & & & L_{\nu\nu} & F_o^\nu
\end{array} \qquad (66)$$

Here the coefficients below the diagonal have been omitted, since in the abridged solution soon to be learned, those below the diagonal are not used. The coefficients are to be read "down to the diagonal, then to the right." The unknowns are the ν Lagrange multipliers.

---

* Maxime Bôcher, <u>Higher Algebra</u> (Macmillan 1907) p. 150.

This type of problem was solved by Gauss,* and is treated satisfactorily in many texts. It arises in geodesy, surveying, and in astronomy, and this accounts for the attentions of Gauss, Bessel, and Encke, who were mainly interested in the problems of adjustment arising in astronomy. Unfortunately they did not give so much attention to problems involving parameters, especially when more than one coordinate is subject to error; and Kummell's paper (1879; see section 1) seems to have been overlooked by modern writers.

It is essential to realize that fundamentally the principle of least squares applies exactly the same whether there are or are not adjustable parameters in the problem. This fact has been woefully overlooked. Critics of least squares have on more than one occasion insisted that least squares was reasonable enough in the problems of adjustment arising in surveying and astronomy, but that it is illogical, equivocal, and unjustified in curve fitting. Statements such as this arise from confusion and misunderstanding, for there is no distinction between the two problems. The student is urged to read the papers by Stewart and Uhler cited earlier (section 1), wherein these things are carefully discussed.

---

* See reference to Gauss in section 9, p. 36.

Example. The plane triangle*

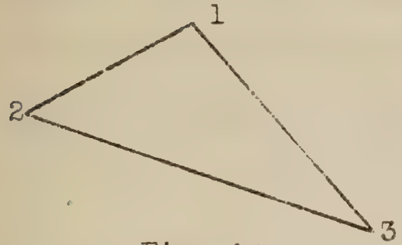

Fig. 4

The angles are measured with a transit.

Observations: X_1, X_2, X_3

Weights: w_1, w_2, w_3

Calc'd values: x_1, x_2, x_3

(The weights might arise from the number of repetitions on each angle).

This is a very simple problem. There is only the one condition

$$x_1 + x_2 + x_3 = 180^\circ \qquad \text{[corresponds to Eq. 50, p. 32]}$$

So here we put

$$F(x_1, x_2, x_3) = x_1 + x_2 + x_3 - 180^\circ$$

(There is only the one condition, so no superscript on the F is needed). This F will be zero when (next page) we are able to insert the adjusted values x_1, x_2, x_3 into it. By inserting the observed values we have

$$F_0 = X_1 + X_2 + X_3 - 180^\circ \qquad \text{[See Eq. 52, p. 33]}$$

F_0 is not zero unless $X_1 + X_2 + X_3$ happens to be exactly 180°, in which case no question of adjustment arises. The derivatives of F are

$$F_1 = F_2 = F_3 = 1 \qquad \text{[See Eq. 51, p. 33]}$$

There is only one L (why?). It could be called L_{11} but no subscript is needed, so we shall use simply L.

$$L = \frac{F_1F_1}{w_1} + \frac{F_2F_2}{w_2} + \frac{F_3F_3}{w_3} = \frac{1}{w_1} + \frac{1}{w_2} + \frac{1}{w_3} \qquad \text{[Eq. 60, p.37]}$$

There is but one normal equation,

$$L\lambda = F_0$$

---

* A more complicated geometric problem, worked out numerically, appears in the next section.

Solution: $\lambda = F_o/L$

The numerator F_o is the amount by which the observed angles fail to close. The denominator L is $\frac{1}{w_1} + \frac{1}{w_2} + \frac{1}{w_3}$ which happens to be equal to $\frac{1}{w_F}$ by Eq. 48, p. 27 (propagation of m.s. error).

After λ is worked out numerically, we find the three residuals by Eq. 58, page 36:

$$V_1 = \frac{1}{w_1}\lambda F_1 = \frac{\lambda}{w_1}$$

$$V_2 = \frac{1}{w_2}\lambda F_2 = \frac{\lambda}{w_2}$$

$$V_3 = \frac{1}{w_3}\lambda F_3 = \frac{\lambda}{w_3}$$

The adjusted angles are then $x_1 = X_1 - V_1$; $x_2 = X_2 - V_2$; $x_3 = X_3 - V_3$.

<u>Note</u> that the sum of the adjusted angles is identically 180°, for

$$x_1 + x_2 + x_3 = X_1 + X_2 + X_3 - (V_1 + V_2 + V_3)$$

$$= X_1 + X_2 + X_3 - \left(\frac{1}{w_1} + \frac{1}{w_2} + \frac{1}{w_3}\right) \frac{X_1 + X_2 + X_3 - 180^\circ}{\frac{1}{w_1} + \frac{1}{w_2} + \frac{1}{w_3}}$$

$$= 180^\circ \text{ exactly}$$

<u>Note</u> that the equations for V_1, V_2, and V_3 are valid no matter how large F_o is. This is a case where there were no higher powers of the residuals to be neglected; cf. notes in section 9 and elsewhere.

<u>Note also</u> that the residuals are inversely proportional to the weights of the observations; that is,

$$V_1 : V_2 : V_3 = \frac{1}{w_1} : \frac{1}{w_2} : \frac{1}{w_3}$$

Thus in this problem the process of least squares simply takes the excess or deficiency F_0 (which will ordinarily be a small amount, perhaps a few minutes of arc) and distributes it among the three angles in inverse proportion to their weights. The student should reflect on this at length. If the action of least squares seems reasonable in this simple problem, it must be so in more complicated ones, even if we are not able to visualize its working so easily. The principle is always the same (the minimizing of $\sum w\,\mathrm{res}^2$ or of χ^2); it is only the conditions to which the adjusted values are subject that differ from one problem to another.

Exercise 1. Show that the condition

$$x_1 + x_2 + x_3 = 180^\circ$$

determines a plane distant $180/\sqrt{3}$ from the origin, and cutting equal intercepts from the axes. The calculated point lies on the plane, and the observed point off it. If the weights are all equal, the distance between the observed and calculated points is to be minimized, in which case the line segment joining the observed and calculated points is perpendicular to the plane $x_1 + x_2 + x_3 = 180^\circ$.

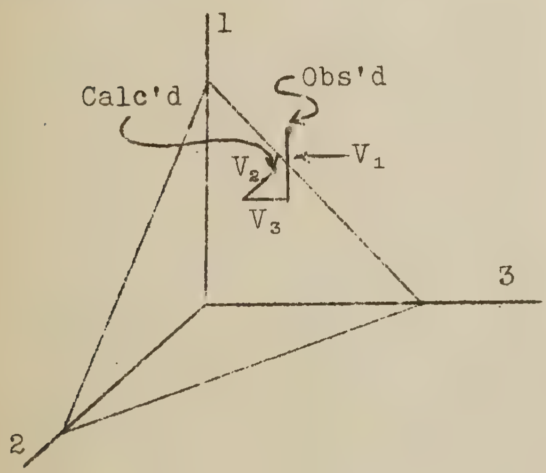

Fig. 5

Exercise 2. All possible plane triangles are represented by points on this plane for which x_1, x_2, and x_3 are positive. Any method of adjustment would consist of picking off some point on this plane, corresponding to a given observed point X_1, X_2, X_3 off the plane.

Exercise 3. If the weights of the observed angles are unequal, the distance between the observed and calculated points is not to be minimized, but rather the quantity

$$w_1(X_1 - x_1)^2 + w_2(X_2 - x_2)^2 + w_3(X_3 - x_3)^2 .$$

Exercise 4. Solve the triangle problem (p. 43) without the Lagrange multiplier.

Hint: Take $\phi^2 = \sum w V^2 = \sum w(X - x)^2$

By the one and only condition on the adjusted values, we may take

$$x_3 = 180^\circ - x_1 - x_2$$

or

$$V_3 = F_0 - V_1 - V_2$$

where, as before, $F_0 = X_1 + X_2 + X_3 - 180^\circ$. Then

$$\phi^2 = w_1V_1^2 + w_2V_2^2 + w_3(F_0 - V_1 - V_2)^2$$

x_1 and x_2 are independent; so are V_1 and V_2. Hence we may set $d\phi^2/dV_1$ and $d\phi^2/dV_2$ both equal to zero. The result is

$$w_1V_1 - w_3(F_0 - V_1 - V_2) = 0$$

$$w_2V_2 - w_3(\quad '' \quad) = 0$$

It follows that $w_1V_1 = w_2V_2$, and that

$$V_1 = (1/w_1)\, F_0 / \{1/w_1 + 1/w_2 + 1/w_3\}$$
$$V_2 = (1/w_2)\, F_0 / \{\quad \text{"} \quad\}$$
$$V_3 = (1/w_3)\, F_0 / \{\quad \text{"} \quad\}$$

which is exactly the same as was obtained earlier with the Lagrange multiplier, page 44.

Note all problems in least squares theoretically can be solved without the use of Lagrange multipliers. Occasionally it may even seem easier to dispense with them, but the truth is that most problems then become hopelessly involved. The elegance and uniformity that they lend to all problems seem to me sufficient to displace all other possible methods in the design of a routine procedure. If Kummell in 1879 had introduced Lagrange multipliers he would surely have accomplished the general solution that he was looking for.

(b) The plane triangle continued. The weights of the adjusted angles and a function of them. Returning to the triangle problem of part (a), suppose we ask for the weight of angle x_1 and of the sum $x_1 + x_2 + x_3$, after adjustment. Of course we know in advance that the weight of this sum must be infinite, since we forced it to be a definite amount, 180°, but it will be interesting to see if this result comes by the routine about to be described. The rules for finding the weights of functions of the adjusted observations are illustrated in what follows, and a more complicated example will be

worked out in the next section. The theoretical proofs will be found in several books on least squares, for example

O. M. Leland Practical Least Squares (McGraw-Hill, 1921)

T. W. Wright and J. F. Hayford Adjustment of Observations (Van Nostrand, 1884, 1906).

Let $G^1 = x_1$ and $G^2 = x_1 + x_2 + x_3$. G^1 and G^2 are the functions whose weights are wanted. As many more functions could be added as desired, but here we shall be content to see just the weights of x_1 and of $x_1 + x_2 + x_3$ worked out. The procedure is as follows. We need to form certain sums, and to this end one can make up the following table, numerical values ordinarily being inserted in place of the symbols in the body of the table.

i	F_i	G_i^1	G_i^2	$F_i/\sqrt{w_i}$	$G_i^1/\sqrt{w_i}$	$G_i^2/\sqrt{w_i}$	Sum
1	1	1	1	$1/\sqrt{w_1}$	$1/\sqrt{w_1}$	$1/\sqrt{w_1}$	for
2	1	0	1	$1/\sqrt{w_2}$	0	$1/\sqrt{w_2}$	numerical
3	1	0	1	$1/\sqrt{w_3}$	0	$1/\sqrt{w_3}$	check

Next step, from columns 5, 6, and 7 form the sums

$$\left[\frac{F_iG_i^1}{w_i}\right] = \frac{1}{w_1}, \qquad \left[\frac{F_iG_i^2}{w_i}\right] = L \quad \text{as defined on page 43.}$$

$$\left[\frac{G_i^1G_i^1}{w_i}\right] = \frac{1}{w_1}, \qquad \left[\frac{G_i^2G_i^2}{w_i}\right] = L \quad \text{" " " " "}$$

[ ] means summation, as in section 9; Gauss' notation.

These sums are appended in the C^1 and C^2 columns, and the solution proceeds as shown under "How obtained".

No.	λ	=	1	C^1	C^2
I	L		F_o	$[F_iG_i^1/w_i] = 1/w_1$	$[F_iG_i^2/w_i] = L$
2			0	$[G_i^1G_i^1/w_i] = 1/w_1$	$[G_i^2G_i^2/w_i] = L$
	How obtained				
3	Eq. I x $(-F_o/L)$		$-F_oF_o/L$	...	...
II	2 + 3		$-F_oF_o/L$	...	...
4	Eq. I x $(-1/w_1L)$			$-1/Lw_1^2$	...
II^1	2 + 4			$1/w_1 - 1/Lw_1^2$	...
5	Eq. I x $(-L/L)$				$-L$
II^2	2 + 5				0

In a numerical solution a Sum column would be used for a check on Eq. II, and the spaces filled in by the ellipses would be filled in with numbers, everything below Eq. 2 being produced according to the directions under "How obtained": for a numerical illustration see pages 64 and 65.

Eq. I gives $\lambda = F_o/L$ as already found at the top of page 44. Looking next at the "1" column in Eq. II we see $-F_oF_o/L$ or $-\lambda F_o$, which by Eq. 65 on page 40 is none other than $-\phi^2$: thus $\Sigma\, wV^2$ is computed in a routine manner without first finding the individual residuals V_1, V_2, and V_3.

The variance coefficient of G^1, or the reciprocal of its weight, appears in the C^1 column of Eq. II^1; and the variance coefficient of G^2, or the reciprocal of its weight, appears in the C^2 column of Eq. II^2.

Before adjustment, the weight of G^1 was w_1, the weight of the observation X_1: after adjustment, the reciprocal of its weight is $1/w_1 - 1/Lw_1^2$.

Now of course

$$\frac{1}{w_1} - \frac{1}{Lw_1^2} < \frac{1}{w_1}$$

which means that the weight of x_1 is greater than the weight of X_1. That is, the weight after adjustment is greater than the weight before, which seems reasonable enough; the observations on the other two angles help to estimate x_1, and to increase our confidence in its value. After the adjustment we feel that we know more about the triangle than before.

In particular if all three angles have the same weight before adjustment, then if $w_1 = w_2 = w_3 = 1$, the weight of x_1 after adjustment is $1/\{1/w_1 - 1/Lw_1^2\} = 1/\{1-1/3\} = 1.5$. In this case the weight of x_1 is 50 percent greater than the weight of X_1; i.e. the adjustment increases the weight by 50 percent; and the same is true for the other angles.

If the weight of an angle has been increased 50 percent, its S.E. has diminished 18 percent, since

$$w_{\text{before}} : w_{\text{after}} = (\text{S.E.}_{\text{after}} : \text{S.E.}_{\text{before}})^2$$ [see Eq. 16 p. 13]

At this point the student will find Art. 129 of Whittaker and Robinson's Calculus of Observations to be enlightening. On page 254 they show that in a quadrilateral, all angles having been measured with equal weight, each adjusted angle has 4/3 the weight of its observed value.

Next consider the weight of $x_1 + x_2 + x_3$. From Eq. I^2 in the form above we see that the reciprocal of its weight is zero; in other words,

the weight of $x_1 + x_2 + x_3$ is infinite; it is therefore known absolutely. The adjustment forced the sum to be 180°, and it is no surprise to find its weight after adjustment to be infinite.

This simple example gives a glimpse of the method for the solution of problems involving rigorous conditions. A guide for systematic computation, and a more complicated example, are given in the next section.

Exercise. Take the values of V_1, V_2, and V_3 found earlier, namely λ/w_1, λ/w_2, and λ/w_3, and show by direct substitution that λF_0, the left-most entry in Eq. II of the tabulation shown above, is actually ϕ^2 or $w_1V_1^2 + w_2V_2^2 + w_3V_3^2$.

12. Geometrical conditions, continued. (a) Systematic computation; the weights of functions of the adjusted values.

Observations: $X_1, X_2, \ldots, X_n$.

Weights: $w_1, w_2, \ldots, w_n$

(Here there will be no subscripts $n + 1$, $n + 2$, ..., $2n$ since there are no y coordinates)

Conditions:

$$F^1(x_1, x_2, \ldots, x_n) = 0$$
$$F^2(x_1, x_2, \ldots, x_n) = 0$$
$$F^3(x_1, x_2, \ldots, x_n) = 0$$
$$F^4(x_1, x_2, \ldots, x_n) = 0$$

(These are Eqs. 50 page 32 except that here there are no parameters).

The 1st step is to write down the conditions, i. e., to select the appropriate F functions; also the G functions whose weights are wanted.* One then works out the values of F_0, which will usually turn out to be small numbers, since the conditions will be nearly but not quite satisfied by the observations; see for instance page 59.

As here indicated, the solution will be illustrated with four conditions; i. e., the number v in Eqs. 50 on p. 32 is taken as 4, which will be the number of Lagrange multipliers (λ). Expansion or contraction to more or fewer conditions is easy. (In the simple triangle problem of section 11, page 43, there was only one condition, and one λ).

We shall assume here that we want to find the weights of two functions of the adjusted values. Let these functions be designated as

---

* By the use of the reciprocal matrix as explained in section 12d, one need not decide on all his G functions at the start; more can be added later without great inconvenience.

$$G^1(x_1, x_2, \ldots, x_n)$$
$$G^2(x_1, x_2, \ldots, x_n)$$

(In the triangle problem of section 11b, G^1 was x_1, and G^2 was $x_1 + x_2 + x_3$; see page 48).

The 2d step is to work out the analytic expressions for the derivatives such as $F_1^1 = \frac{dF^1}{dx_1}$, $F_3^2 = \frac{dF^2}{dx_3}$, etc.; also $G_1^1 = \frac{dG^1}{dx_1}$, $G_4^2 = \frac{dG^2}{dx_4}$, etc.; see, for instance, pages 60 and 61.

The 3d step is to work out the numerical values of all these derivatives; see, for instance, page 61. In each case, the observed values $X_1, X_2, \ldots, X_n$ are used in place of $x_1, x_2, \ldots, x_n$, since approximate values of the derivatives are usually close enough; at least they will have to suffice till we can get better ones. The following table is made up, numerical values being inserted in the spaces. Naturally, more or fewer columns will be needed in various problems, and different computers will work differently even on the same problem. For example, one might wish to record w_i in the 2d column and put $\sqrt{w_i}$ in the 3d. Only general directions can be given in advance of a specific problem.

Table 1

i	$\sqrt{w_i}$	F_i^1	F_i^2	F_i^3	F_i^4	G_i^1	G_i^2
1	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	-	-	-	-	-	-	-

The 4th step is to form table 2, derived from table 1 by dividing the F and G derivatives by the corresponding $\sqrt{w_i}$ as indicated in the headings of table 2.

Table 2.--The matrix for the formation of the normal equations.*

i	$\frac{F_i^1}{\sqrt{w_i}}$	$\frac{F_i^2}{\sqrt{w_i}}$	$\frac{F_i^3}{\sqrt{w_i}}$	$\frac{F_i^4}{\sqrt{w_i}}$	$\frac{G_i^1}{\sqrt{w_i}}$	$\frac{G_i^2}{\sqrt{w_i}}$	Sum S_i
1	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-
:	:	:	:	:	:	:	:
n	-	-	-	-	-	-	-
Sum	-	-	-	-	-	-	-✓

The sums across the bottom and at the right are used for checking the formation of the normal equations. With a machine having two multiplier registers, one for cumulating the quotients, and the other for reading individual quotients, one may cumulate one set of sums without extra effort; see also page 101.

5th step. The normal equations 67, together with the C^1 and C^2 columns, are now made up by summing squares and cross-products from Table 2. Thus, L_{11} of Eqs. 66 is the sum of the squares of the n entries in the $F_i^1/\sqrt{w_i}$ column; L_{12} is the sum of the n cross-products in the $F_i^1/\sqrt{w_i}$ and $F_i^2/\sqrt{w_i}$ columns (refer back to Eq. 60, p. 37). Instead of writing L_{11}, L_{12}, etc., as was done in Eqs. 66, it is better now to fill in with the symbols that refer back directly to table 2. [ ] denotes summation, as before (Gauss' notation).

---

* Concerning the use of the term matrix here, see the footnote on p. 161.

Normal Eqs.

No.	λ_1	λ_2	λ_3	λ_4	=	1	C^1	C^2	Sum
I	$\left[\frac{F_i^1F_i^1}{w_i}\right]$	$\left[\frac{F_i^1F_i^2}{w_i}\right]$	$\left[\frac{F_i^1F_i^3}{w_i}\right]$	$\left[\frac{F_i^1F_i^4}{w_i}\right]$	=	F_o^1	$\left[\frac{F_i^1G_i^1}{w_i}\right]$	$\left[\frac{F_i^1G_i^2}{w_i}\right]$	...
2		$\left[\frac{F_i^2F_i^2}{w_i}\right]$	$\left[\frac{F_i^2F_i^3}{w_i}\right]$	$\left[\frac{F_i^2F_i^4}{w_i}\right]$	=	F_o^2	$\left[\frac{F_i^2G_i^1}{w_i}\right]$	$\left[\frac{F_i^2G_i^2}{w_i}\right]$	...
3			$\left[\frac{F_i^3F_i^3}{w_i}\right]$	$\left[\frac{F_i^3F_i^4}{w_i}\right]$	=	F_o^3	$\left[\frac{F_i^3G_i^1}{w_i}\right]$	$\left[\frac{F_i^3G_i^2}{w_i}\right]$	Sum column used for check
4				$\left[\frac{F_i^4F_i^4}{w_i}\right]$	=	F_o^4	$\left[\frac{F_i^4G_i^1}{w_i}\right]$	$\left[\frac{F_i^4G_i^2}{w_i}\right]$	
5						0	$\left[\frac{G_i^1G_i^1}{w_i}\right]$	$\left[\frac{G_i^2G_i^2}{w_i}\right]$	...

No entries below the diagonal because of symmetry.

(67)

The Sum column at the right checks the formation of the normal equations. Herein are entered (in pencil) the cumulation of the cross multiplications formed with the Sum column of table 2; these should agree with the sums of the terms in the normal equations, the "1" column excluded; see table 3 in part (b), and the check formed immediately below. If no errors are found, the sums entered in pencil at the right of the normal equations are altered to include the "1" column, and the solution proceeds, being checked at the pivotal points (see the check marks in Eqs. II, III, and IV of the numerical solution in part b). The sums $[G_i^1G_i^1/w_i]$ and $[G_i^2G_i^2/w_i]$ must be checked otherwise, as by repetition.

The 0 in the bottom row of the normal equations is appended for the computation of ϕ^2; the columns C^1 and C^2 for the computation of the weights of the functions G^1 and G^2.

The solution of the equations is carried out by the routine process already seen in simplified form on page 49, and to be illustrated more fully on pages 64-65, and symbolically on page 107. When the numerical values of the Lagrange multipliers (λ) have been worked out, the residuals $V_1, \ldots, V_n$ are calculated by Eq. 58 and then used to find the "adjusted observations" $x_1, x_2, \ldots, x_n$ as follows:

$$\left.\begin{aligned}
x_1 &= X_1 - V_1 = X_1 - \frac{1}{w_1}(\lambda_1 F_1^1 + \lambda_2 F_1^2 + \lambda_3 F_1^3 + \lambda_4 F_1^4) \\
x_2 &= X_2 - V_2 = X_2 - \frac{1}{w_2}(\lambda_1 F_2^1 + \lambda_2 F_2^2 + \lambda_3 F_2^3 + \lambda_4 F_2^4) \\
&\vdots \quad \vdots \\
x_n &= X_n - V_n = X_n - \frac{1}{w_n}(\lambda_1 F_n^1 + \lambda_2 F_n^2 + \lambda_3 F_n^3 + \lambda_4 F_n^4)
\end{aligned}\right\} \quad (68)$$

(b) Numerical example--a surveying problem. A surveying party measures the sides and angles of the plane triangle P Q R, with the following results:

On angle P:	51° 06'
	08
	05
	06
Average	51° 06'.25
On angle Q:	95° 05'
	04
Average	95° 04'.5
On angle R:	33° 49'
	50
	33° 49'.5

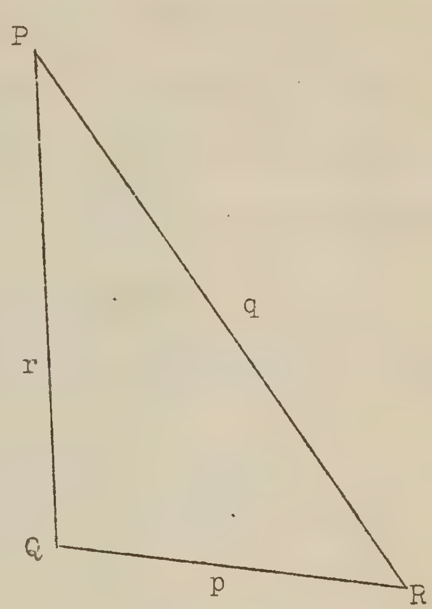

Fig. 6

Side p: 1723.7 ft.

" q: 2205.4 "

" r: 1232.7 "

The transit man, from previous experience, has reason to believe that the r.m.s. error of single measurements on one angle is about one minute of arc, or .00029 radians. He takes the r.m.s. error of the chainmen to be one foot in 10,000 feet, and in proportion to the square root of the distance chained. The weights of the observations on the angles and sides are then in ratios as follows:

$$w_P : w_Q : w_R : w_p : w_q : w_r = 4/0.000\,29^2 : 2/0.000\,29^2$$
$$: 2/0.000\,29^2 : 10000/1724 : 10000/2205 : 10000/1233$$

Looking back at Eq. 13 on page 12 we see that what has just been applied is the fact that w_f is inversely proportional to the variance of f. The factor of proportionality is σ^2, the mean square error of observations of unit weight. Since weights are relative and not absolute, this factor of proportionality (σ^2) is arbitrary and can be chosen for convenience; accordingly we let

$$\sigma^2 = \frac{1}{2}(0.000\ 29)^2 = 4.23 \times 10^{-8}$$

whereupon the weights take these simple values:

$$w_P = 2, \qquad w_Q = 1, \qquad w_R = 1,$$

$$w_p = 24.6 \times 10^{-8}, \quad w_q = 19.2 \times 10^{-8}, \quad w_r = 34.3 \times 10^{-8}$$

It should be noted that the final adjusted values of the sides and angles, also their S.Es., are in no way dependent on the arbitrary choice made for σ^2 ; if σ^2 is doubled, all the weights are also doubled, and the S.Es. of all functions are left unaltered. Likewise χ^2 is unaltered.

<u>1st step</u>. The adjustment must be carried out to enforce the following three geometrical

<u>Conditions</u>

$$\frac{\sin P}{p} = \frac{\sin Q}{q} = \frac{\sin R}{r}$$

$$P + Q + R = 180^\circ + \varepsilon$$

ε being the spherical excess, which, owing to the small size of the triangle, will here be taken as zero. If it were other than zero, F_0^3 (below) would be altered by the amount ε, and the adjusted values of the sides and angles and their S.Es. would all be affected in an obvious manner.

For forcing the three conditions, let us set

$$F^1\ (P, Q, R, p, q, r) = (\sin P)/p - (\sin Q)/q$$

$$F^2\ (\text{" " " " " "}) = (\sin P)/p - (\sin R)/r$$

$$F^3\ (\text{" " " " " "}) = P + Q + R - 180^\circ$$

(The number v of Eqs. 50 is here equal to 3)

The observed values of P, Q, R, p, q, r, when substituted into F^1, F^2, F^3 do not give zeros, but give the small numbers F^1_o, F^2_o, F^3_o, which by direct substitution are found to be:

$$F^1_o = \sin 51^\circ\ 06'.25/1723.7 - \sin 95^\circ\ 04'.5/2205.4$$

$$= -1.3271 \times 10^{-7}$$

$$F^2_o = \sin 51^\circ\ 06'.25/1723.7 - \sin 33^\circ\ 49'.5/1232.7$$

$$= -0.5416 \times 10^{-7}$$

$$F^3_o = 51^\circ\ 06'.25 + 95^\circ\ 04'.5 + 33^\circ\ 49'.5 - 180^\circ$$

$$= 0^\circ\ \ 0'.25 = 7.27 \times 10^{-5} \text{ radians}$$

If it had happened that the observations satisfied the conditions exactly, then F^1_o, F^2_o, and F^3_o would have turned out to be zeros, and the adjusted values would have been identical with those observed. As it is, the observations satisfy the conditions nearly but not exactly, i.e., F^1_o, F^2_o, and F^3_o are small but not zeros.

F^3_o is the amount by which the sum of the angles exceeds 180°. In the simpler problem wherein the sides were not measured (vide supra, section 11a) it turned out that the least squares adjustment was simply an apportionment of this discrepancy among the three angles in inverse

proportion to their weights. Now, however, the sides are involved, wherefore the adjustment, though as reasonable as before, will not be so easy to arrive at. By looking ahead at page 67 we see that, in contrast with the residuals on page 44, the adjustments on the angles will not now be all in the same direction.

Now suppose that for some reason or other we should like to know the weights of

Angle P

The sum P + Q + R

The area of the triangle, which may be expressed as $\frac{1}{2}$ pr sin Q

Any number of others could be added (at increased labor) but three will suffice here. For those just named we take the three G functions

$$G^1 = P, \qquad G^2 = P + Q + R, \qquad G^3 = \frac{1}{2} pr \sin Q$$

2d step. The derivatives of the F functions are:

$F^1_P = \cos P/p$	$F^2_P = \cos P/p$	$F^3_P = 1$
$F^1_Q = -\cos Q/q$	$F^2_Q = 0$	$F^3_Q = 1$
$F^1_R = 0$	$F^2_R = -\cos R/r$	$F^3_R = 1$
$F^1_p = -\sin P/p^2$	$F^2_p = -\sin P/p^2$	$F^3_p = 0$
$F^1_q = \sin Q/q^2$	$F^2_q = 0$	$F^3_q = 0$
$F^1_r = 0$	$F^2_r = \sin R/r^2$	$F^3_r = 0$

The derivatives of the G functions are:

$G_P^1 = 1$	$G_P^2 = 1$	$G_P^3 = 0$
The other	$G_Q^2 = 1$	$G_Q^3 = \frac{1}{2}\, pr \cos Q$
five are	$G_R^2 = 1$	$G_R^3 = 0$
zero	The other	$G_p^3 = \frac{1}{2}\, r \sin Q$
	three are	$G_q^3 = 0$
	zero	$G_r^3 = \frac{1}{2}\, p \sin Q$

3d step. The nearest numerical approximations that we can produce for these derivatives are found by substituting the observed angles and sides into the expressions just worked out, and these will be more than close enough.

Table 1.--The derivatives (3d step).

i	w_i	$\sqrt{w_i}$	$10^6 F_i^1$	$10^6 F_i^2$	F_i^3	G_i^1	G_i^2	G_i^3
P	2	1.41	364	364	1	1	1	0
Q	1	1	40.1	0	1	0	1	-93916
R	1	1	0	-674	1	0	1	0
p	$24.6 \cdot 10^{-8}$	$4.96 \cdot 10^{-4}$	-.262	-.262	0	0	0	613.9
q	19.2 "	4.38 "	.205	0	0	0	0	0
r	34.3 "	5.86 "	0	.366	0	0	0	858.5

4th step $\sqrt{w}$ is now used as a divisor to form table 2 from table 1.

Table 2.--The matrix for the formation of the normal equations.

i	$F^1_i\sqrt{w_i}10^3$	$F^2_i\sqrt{w_i}10^3$	$F^3_i\sqrt{w_i}$	$G^1_i\sqrt{w_i}$	$G^2_i\sqrt{w_i}$	$G^3_i\sqrt{w_i}10^{-6}$	Sum
P	0.257	0.257	0.707	0.707	0.707	0	2.635
Q	.040	0	1	0	1	-.094	1.946
R	0	-.674	1	0	1	0	1.326
p	-.528	-.528	0	0	0	1.238	0.182
q	.468	0	0	0	0	0	.468
r	0	.625	0	0	0	1.465	2.090
Sum	.237	-.320	2.707	.707	2.707	2.609	8.647✓

The powers of 10 in table 2 are chosen with regard to convenience, and to bring the number of decimals to uniformity from column to column, to facilitate the cumulation of squares and cross-products in forming the normal equations (the next step). At this stage one may also cut off superfluous figures, reserving, as a rule, not more than three or four in the largest number occurring in any one column. This often means that some other entries in the same column appear as zeros, but this is as it should be.

<u>5th step</u>. The cumulations of squares and cross-products from the columns of table 2 provide the coefficients required for the normal equations 67 page 55. For instance*

$$10^6\left[\frac{F^1F^1}{w}\right] = 0.257^2 + .040^2 + 0^2 + .528^2 + .468^2 + 0^2 = 0.565$$

---

* The subscript i will usually be omitted for convenience hereafter.

as seen under λ_1 in the normal equations. Also

$$10^3 \left[\frac{F^2F^3}{w}\right] \doteq .257 - .707 + 0 - .674 + 0 + 0 + 0 + 0 = -0.492$$

as seen under λ_3. The student should verify the whole set appearing in table 3.

Table 3.--The cumulation of squares and cross products from table 2 for the formation of the normal equations.

$10^6 \left[\frac{F^1F^1}{w}\right] = 0.565,$	$10^6 \left[\frac{F^1F^2}{w}\right] = 0.345,$	$10^3 \left[\frac{F^1F^3}{w}\right] = 0.222$
	$10^6 \left[\frac{F^2F^2}{w}\right] = 1.190,$	$10^3 \left[\frac{F^2F^3}{w}\right] = -.492$
		$\left[\frac{F^3F^3}{w}\right] = 2.500$
$10^3 \left[\frac{F^1G^1}{w}\right] = 0.182,$	$10^3 \left[\frac{F^1G^2}{w}\right] = 0.222,$	$10^{-3} \left[\frac{F^1G^3}{w}\right] = -0.657$
$10^3 \left[\frac{F^2G^1}{w}\right] = 0.182,$	$10^3 \left[\frac{F^2G^2}{w}\right] = -0.492,$	$10^{-3} \left[\frac{F^2G^3}{w}\right] = 0.262$
$\left[\frac{F^3G^1}{w}\right] = 0.500,$	$\left[\frac{F^3G^2}{w}\right] = 2.500,$	$10^{-6} \left[\frac{F^3G^3}{w}\right] = -0.094$
$\left[\frac{G^1G^1}{w}\right] = 0.500,$	$\left[\frac{G^2G^2}{w}\right] = 2.500,$	$10^{-12} \left[\frac{G_3G_3}{w}\right] = 3.687$
		$\left[\frac{F^1S}{w}\right] = .878^*$
		$\left[\frac{F^2S}{w}\right] = .994^*$
		$\left[\frac{F^3S}{w}\right] = 5.135^*$

Check*:

$$.565 + .345 + .222 + .182 + .222 - .657 = .879$$

$$.345 + 1.190 - .492 + .182 - .492 + .262 = .995$$

$$.222 - .492 + 2.500 + .500 + 2.500 - .094 = 5.136$$

---

* Powers of 10 disregarded in sum checks.

Combined solution of the normal equations, the computation

No.	Factors	$10^{-6}\lambda_1$	$10^{-6}\lambda_2$	$10^{-3}\lambda_3$	= 1
I		.565	.345	.222	$-.133\times10^{-6}$
2			1.190	-.492	-.054
3				2.500	.073
4					0
5	.345/ .565 = .6106		-.211	-.136	.081
II			.979	-.628	-.027
6	.222/ .565 = .3929			-.087	.052
7	.628/ .979 = .6415			-.403	.017
III				2.010	.142
8	.133/ .565 = .2354				-.031
9	.027/ .979 = .0276				-.001
10	.142/2.010 = .0706				-.010
IV					-.042
13				$10^{-6}\lambda_1$ =	$-.308\times10^{-6}$
12				$10^{-6}\lambda_2$ =	$.073\times10^{-6}$
11				$10^{-3}\lambda_3$ =	$.071\times10^{-6}$
14	.182/ .565 = .3221				
15	.071/ .979 = .0725				
16	.474/2.010 = .2358				
IV^1					
17	.222/ .565 = .3929				
18	.628/ .979 = .6415				
19	2.010/2.010 = 1				
IV^2					
20	.657/ .565 = 1.1623				
21	.663/ .979 = .6772				
22	.589/2.010 = .2930				
IV^3					

(The powers of 10 written at the tops of the "1", C^1, C^2, and C^3 columns are understood to apply all the way down)

of $\sum wV^2$, and the weights of three functions

C^1	C^2	C^3	Sum	How obtained
$.182\times10^{-3}$	$.222\times10^{-3}$	$-.657\times10^{3}$	0.746	
.182	-.492	.262	0.941	
.500	2.500	-.094	5.209	
.500	2.500	3.687	6.573	
-.111	-.136	.401	-.456	I(-.6106)
.071	-.628	.663	.485	(2) + (5)
-.072	-.087	.258	-.293	I(-.3929)
.046	-.403	.425	.311	II(+.6415)
.474	2.010	.589	5.227	(3) + (6) + (7)
.043	.052	-.155	.176	I(-.2354)
-.002	-.017	-.018	-.013	II(+.0276)
-.033	-.142	-.042	-.369	III(-.0020)
.508	2.427	3.472	6.367	(4) + (8) + (9) + (10)
				Subst. from 11 & 12 into I
				" " 11 into II
				III ÷ 2.010
-.059				I(-.3221)
-.005				II(-.0725)
-.112				III(-.2358)
.324				(4) + (14) + (15) + (16)
	-.087			I(-.3929)
	-.403			II(+.6415)
	-2.010			III(-1)
	0.000			(4) + (17) + (18) + (19)
		.764		I(-1.1623)
		-.449		II(-.6772)
		-.173		III(-.2930)
		3.829		(4) + (20) + (21) + (22)

The sums at the bottom of page 63 do not tally exactly with the numbers $[F^1S/w] = 0.878$, $[F^2S/w] = 0.994$, and $[F^3S/w] = 5.135$ seen just above, but the agreement is within errors of rounding off, whereupon we conclude that the arithmetic in table 3 is correct, save for the three [GG/w] sums, which must be checked independently, as by repetition in reverse order. The cumulations shown in table 3 are then entered into Eqs. I, 2, 3, 4 of the tabular scheme for the normal equations on the two preceding pages. The numbers entered in the "1" column of Eqs. I, 2, and 3 come from the values of F_0^1, F_0^2, and F_0^3 on page 59 after multiplication by appropriate powers of 10 to produce decimals of the same denomination as the other parts of the normal equations. (The factor 10^{-6} applies to the whole of the "1" column).

The sums at the right of the normal equations are not the numbers .879, .995, and 5.136 previously seen in the check under table 3 but are these numbers to which have been added the corresponding entries of the "1" column; the normal equations thus start off with a Sum column that provides checks at the pivotal points of the solution (note the check marks in Eqs. II, III, and IV).

The solution proceeds according to the directions under "How obtained". The same system of solution has been seen in simple problems on pages 10, 24, and 49, and will be seen again on page 107 and in section 20.

(c) Conclusions from the solution of the normal equations

1° From Eq. IV, ϕ^2 or $\Sigma wV^2 = 0.042 \cdot 10^{-6}$

It follows from Eq. 21 on page 19 that

$$\sigma^2(\text{ext}) = 0.042 \cdot 10^{-6} \div (6-3) = 1.4 \cdot 10^{-8}$$

Since this is only about one-third the prior σ^2 arbitrarily chosen on page 58, we conclude that so far there is no indication of blunders in the observations or recording.

2° Eqs. 13, 12, and 11 in the solution on pages 64 and 65 give

$$\lambda_1 = -0.308, \quad \lambda_2 = 0.073, \quad \lambda_3 = 0.071 \cdot 10^{-3}$$

These used in Eq. 58 page 36 give

$$\begin{aligned}
V_P &= \tfrac{1}{2}(\lambda_1 F_P^1 + \lambda_2 F_P^2 + \lambda_3 F_P^3) = -0.0000075 \text{ radian} = -0.03 \text{ min.} \\
V_Q &= \lambda_1 F_Q^1 + \lambda_2 F_Q^2 + \lambda_3 F_Q^3 = .0000582 \quad '' \quad = \quad .20 \quad '' \\
V_R &= \lambda_1 F_R^1 + \lambda_2 F_R^2 + \lambda_3 F_R^3 = .0000214 \quad '' \quad = \quad .07 \quad ''
\end{aligned}$$

$$\begin{aligned}
V_p &= (10^8/24.6)(\lambda_1 F_p^1 + \lambda_2 F_p^2 + \lambda_3 F_p^3) = 0.25 \text{ ft.} \\
V_q &= (10^8/19.2)(\lambda_1 F_q^1 + \lambda_2 F_q^2 + \lambda_3 F_q^3) = -.33 \quad '' \\
V_r &= (10^8/34.3)(\lambda_1 F_r^1 + \lambda_2 F_r^2 + \lambda_3 F_r^3) = .08 \quad ''
\end{aligned}$$

for the six residuals. It is important to note that the derivatives required here are already worked out in table 1, page 61.

3° By using these residuals with Eq. 63 page 38, we find that the adjusted value of

Angle P is 51° 06'.25 + 0'.03 = 51° 06'.28

Angle Q is 95° 04'.5 - .20 = 95° 04'.30

Angle R is 33° 49'.5 - .07 = 33° 49'.43

Side p′ is 1723.7 - 0.25 = 1723.45 ft.

Side q is 2205.4 + .33 = 2205.73 "

Side r is 1232.7 - .08 = 1232.62 "

Remark Perfect closure (3d condition on page 59) may be secured by lowering angle R by the trifling amount 0'.01; the value 33° 49'.42 so obtained, along with the other adjusted angles and sides just written, will satisfy also the 1st and 2d conditions on page 59 to within 1 part in $\frac{1}{2}$ million, which is about all we should ask for. Whenever, as happened here, one or more of the conditions fails owing to cumulated inexactness of rounding off, the computer is at liberty to manipulate the terminal figure of one or more of the residuals, raising or lowering it a unit or so to force the conditions. If not inconvenient, he will ordinarily (as was just done here) select the quantities of least weight for any such manipulations. The amount involved will be small compared with the S. Es. of the final results. (Cf. also page 166).

4° The weights and S. Es. of the three G functions (page 60) are found as follows.

From Eq. IV^1 the weight of the adjusted angle P is $1/0.324 \cdot 10^{-6} \cdot 10^{+6}$ = 1/0.324. In other words, 0.324 is the variance coefficient of angle P. Then with $\sigma^2 = 4.23 \cdot 10^{-8}$ (page 58), it turns out that the S. E. of the adjusted angle P is $(4.23 \cdot 10^{-8} \cdot 0.324)^{\frac{1}{2}} = 1.2 \cdot 10^{-4}$ radians = 0.40 min. So

Angle P is 51° 06'.3 ± 0'.4

From Eq. IV^2 the weight of the adjusted sum of P + Q + R is 1/0 or ∞, as predicted. Hence the sum of the adjusted angles would be written

$P + Q + R = 180°$ absolutely

From IV^3 the weight of the area $\frac{1}{2}$ pq sin R is $1/2.302 \times 10^{12}$. Its S.E. is therefore $(4.23 \times 10^{-8} \times 2.30 \times 10^{12})^{\frac{1}{2}} = 312$ square feet; therefore the adjusted value of

The area is 1058028 ± 312 sq. ft.

The area would better be written (105803 ± 31) x 10 square feet, since not more than two figures of the S.E. could be assumed known. In acres,

The area = 24.2890 ± 0.0072 acres

(The area is found by using the adjusted values of p, q, and R and taking $\frac{1}{2}$ pq sin R. Of course one could as well use $\frac{1}{2}$ qr sin P or $\frac{1}{2}$ pr sin Q for the area; one is as good as another).

Exercise 1. Prove by Eq. 48 section 8 that after adjustment the weight of the area is a little more than double its weight before adjustment.

Hint: By using Eq. 48 p. 27 we find that

$$
\begin{aligned}
1/w_{area} &= area^2 \{1/p^2w_p + 1/q^2w_q + (\cot^2 R)/w_R\} \\
&= 1.12 \times 10^{12} \{1.37 + 1.07 + 2.24\} \\
&= 5.25 \times 10^{12} \quad \text{before adjustment} \\
\therefore w_{area} &= 0.19 \times 10^{-12} \quad \text{(before)}
\end{aligned}
$$

We had w_{area} = weight of G^3 = 0.43×10^{-12} (after)

The stated result follows at once.

Exercise 2. (From L. D. Weld's Theory of Errors and Least Squares, Macmillan, 1916). Take the line AB, on which are located points C and D. The whole line and its segments are measured with the same rule under similar conditions, the results being

X_1 = AC = 45.10	cm.,	mean	of	2	observations
X_2 = AD = 77.96	"	"	"	3	"
X_3 = CD = 32.95	"	"	"	2	"
X_4 = CB = 98.36	"	"	"	3	"
X_5 = DB = 65.55	"	"	"	2	"
X_6 = AB =143.55	"	"	"	4	"

Problem: Find the least squares values of the lengths.

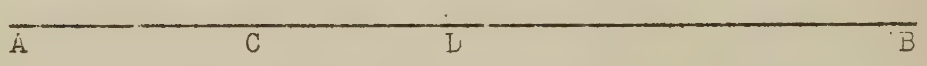

Fig. 7

Take $w_1 = 2$, $w_2 = 3$, $w_3 = 2$, $w_4 = 3$, $w_5 = 2$, $w_6 = 4$.

Conditions:
$$F^1 = x_1 + x_3 + x_5 - x_6 = 0$$
$$F^2 = x_1 - x_2 + x_3 = 0$$
$$F^3 = x_3 - x_4 + x_5 = 0$$

Show that the normal equations are as follows:

No.	λ_1	λ_2	λ_3	= 1	Sum
I	21	12	12	$60 \cdot 10^{-2}$	105
2		16	6	108 "	142
3			16	168 "	202
4				0	328

Solution: $\lambda_1 = .1145$, $\lambda_2 = -.0952$, $\lambda_3 = -.1552$

Residuals: (by applying Eq. 58)

$V_1 = -0.0096$

$V_2 = -\ .0317$

$V_3 = +\ .0680$

$V_4 = -\ .0517$

$V_5 = +\ .0203$

$V_6 = +\ .0286$

Adjusted values: AC = 45.110 cm.

AD = 77.992

CD = 32.882

CB = 98.412

DB = 65.530

AB =143.522

(AB actually turns out to be 143.521 cm., but the last decimal is raised one unit to satisfy the first condition. The other two conditions are satisfied perfectly by the adjusted segments).

Exercise 3. By Eq. IV in the solution of the normal equations of the preceding exercise, the minimized value of $\sum wV^2$ is 0.0246.

Exercise 4. Find the S. Es. of AB and AD, taking the S. E. σ of a single measurement to be 0.05 cm.

Exercise 5. (a) Show that the estimate of σ made from $\sum wV^2$ is $\sigma(\text{ext}) = 0.0$
(b) Show that with $\sigma = 0.05$, χ^2 = about 10, and $P(\chi^2) = 0.02$, wherefore we might say that the discordance between the observed lengths of the segments is somewhat larger than one might expect from previous experience. Note: Since the individual measurements were not recorded, there is no possibility of estimating σ from the original observations; i.e. we have no $\sigma(\text{int})$ to compare with the prior σ and $\sigma(\text{ext})$.

Exercise 6. The three inside edges of a parallelopiped are measured with calipers and a linear scale; and the volume is measured in cubic units by filling it with mercury, which is afterward poured into a graduated cylinder. The results of a set of observations are as follows:

	Mean	n	S. D.
On edges parallel to the x-direction,	X_1 (cm.)	n_1	s_1
On edges " " " y-direction,	X_2 "	n_2	s_2
On edges " " " z-direction,	X_3 "	n_3	s_3
On the volume,	X_4 (cc.)	n_4	s_4

It seems safe to pool the S. Ds. on the linear measurements to get an estimate of their S. Es. or weights, and if $n_1 + n_2 + n_3$ is fairly large (20 or 30), this estimate may be fairly reliable. For as reliable an estimate on the S. E. of the direct determinations of volume, n_4 itself would have to be as large as 20 or 30. Assuming that we have these numbers, or have gotten estimates by previous experience, we assign weights as follows:

$$w_1 = n_1\sigma^2/\sigma_1^2$$

$$w_2 = n_2\sigma^2/\sigma_1^2$$

$$w_3 = n_3\sigma^2/\sigma_1^2$$

$$w_4 = n_4\sigma^2/\sigma_4^2$$

σ_1 being the S. E. of a single linear measurement, and σ_4 that of a single volume measurement by mercury. σ^2, as in section 5, is an arbitrary factor of proportionality, the S. E. of observations of unit weight. If it is set equal to σ_1^2, we should have the convenient system of weights,

$$w_1 :: w_2 : w_3 : w_4 = n_1 : n_2 : n_3 : n_4\sigma_1^2/\sigma_4^2$$

The weights having been settled on, we can proceed. The one and only condition on the adjusted values is that

$$x_4 = x_1x_2x_3$$

whence we put $F = x_4 - x_1x_2x_3$

Suppose we need the S. E. of the volume after adjustment; we set

$$G = x_4$$

(a) Show that the one and only normal equation is $L\lambda = F_o$, whence $\lambda = F_o/L$, where

$$L = X_4^2 \{1/X_1^2w_1 + 1/X_2^2w_2 + 1/X_3^2w_3 + 1/X_4^2w_4\}$$

(b) In tabular form, the normal equation for finding λ, ϕ^2, and the weight of the adjusted volume, is as follows:

No.	λ	1	C
I	L	F_o	$1/w_4$
2		0	$1/w_4$
II		$-F_o\lambda$	$\frac{1}{w_4} - \frac{1}{w_4L}$

(c) The weight of the adjusted volume is $(1/w_4)(1 - 1/L)$

(d) (The S. E. of the adjusted volume)^2 = $\sigma^2(1/w_4)(1 - 1/L)$
Note that this is smaller than the S. E. of the volume before adjustment by the fractional amount $1/L$.

(e) The minimized sum of the weighted squares of the residuals, ϕ^2, is $F_o\lambda$.

(f) The estimate of σ^2 by external consistency (section 6c) is

$$\sigma^2(\text{ext}) = F_o\lambda/3$$

(g) What would you say if σ^2(ext) were much larger than σ^2, i.e. $P(\chi)$ small? Suggestions: Edges not parallel; lack of perpendicularity; measurements not so good as initially supposed (i.e. σ_1 or σ_4 too small); just happened to be so.

(h) Show that after adjustment the S. E. of the first edge is

$$\sigma\sqrt{(1/w_1)(1 - X_4/X_1L)}$$

(d) Shorter method of computing the weights of a large number of functions*. The theory on which the weights of the three G functions were calculated in parts (b) and (c) rests on the fact that**

$$1/(\text{wt. of G}) = \left[\frac{GG}{w}\right] - \left[\frac{F^1G}{w}\right]B' - \left[\frac{F^2G}{w}\right]B'' - \left[\frac{F^3G}{w}\right]B''' \qquad (69)$$

where B', B'', and B''' satisfy the equations

$$\left.\begin{aligned} L_{11}\, B' + L_{12}\, B'' + L_{13}\, B''' &= \left[\frac{F^1G}{w}\right] \\ L_{21}\, B' + L_{22}\, B'' + L_{23}\, B''' &= \left[\frac{F^2G}{w}\right] \\ L_{31}\, B' + L_{32}\, B'' + L_{33}\, B''' &= \left[\frac{F^3G}{w}\right] \end{aligned}\right\} \qquad (70)$$

In other words, the auxiliary constants B', B'', and B''' satisfy the normal equations 67 p. 55 with the "C" column

$$\left[\frac{F^1G}{w}\right]$$

$$\left[\frac{F^2G}{w}\right]$$

$$\left[\frac{F^3G}{w}\right]$$

replacing the "1" column..

One may, if he chooses, solve for the Lagrange multipliers and any set of auxiliary constants B', B'', B''' by first of all calculating the reciprocal matrix

---

* To be omitted on first reading; the suggestion is that the reader return to this after a study extending through section 17.

** Gauss, Theoria Combinationis (cited in section 6c) Art. 29.

$$A^{-1} = \begin{vmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{vmatrix}$$ (see Exs. 2, 4, and 5 of section 17)

and then using it to calculate the Lagrange multipliers and the auxiliary multipliers as follows:

$$\left.\begin{aligned} \lambda_1 &= F_0^1 c_{11} + F_0^2 c_{12} + F_0^3 c_{13} \\ \lambda_2 &= F_0^1 c_{21} + F_0^2 c_{22} + F_0^3 c_{23} \\ \lambda_3 &= F_0^1 c_{31} + F_0^2 c_{32} + F_0^3 c_{33} \end{aligned}\right\} \qquad (71)$$

$$\left.\begin{aligned} B' &= \left[\frac{F^1G}{w}\right] c_{11} + \left[\frac{F^2G}{w}\right] c_{12} + \left[\frac{F^3G}{w}\right] c_{13} \\ B'' &= \left[\;''\;\right] c_{21} + \left[\;''\;\right] c_{22} + \left[\;''\;\right] c_{23} \\ B''' &= \left[\;''\;\right] c_{31} + \left[\;''\;\right] c_{32} + \left[\;''\;\right] c_{33} \end{aligned}\right\} \qquad (72)$$

The Lagrange multipliers (λ), after being calculated from Eqs. 71, are used in Eqs. 58 page 36 to compute the residuals $V_1, \ldots, V_n$, just as was done on page 67. The auxiliary constants B', B'', and B''' from Eqs. 72 are used in Eq. 69 to find the weight of the function G. It will be noticed that the coefficients multiplying the c's in Eqs. 71 would already have been worked out by the first step outlined on page 52 and carried out numerically on page 59. The coefficients in Eqs. 72 come by cumulating squares and cross products from table 2 of the fourth step, page 54, as was carried out numerically in table 3 on page 63. It is not difficult to extend tables 2 and 3 to care for a new G function any time one is desired.

The work then proceeds rapidly, the reciprocal matrix A^{-1} being used over and over in Eqs. 72 for all the G functions. If one is working with a fairly good sized number of G functions, this scheme will undoubtedly

save considerable time over the direct computation illustrated in section 12b.

A distinct advantage of using the auxiliary multipliers is that the reciprocal matrix, once computed, is ready for use any time a new C column is produced, whereas with the direct solution in section 12b it is no little trouble to introduce a new C column after a solution has once been carried through.

The three G functions used in section 12b will serve for an illustration. To calculate the reciprocal matrix A^{-1} we take the coefficients of the unknowns in the normal equations on pages 64 and 65, and put the unit matrix on the right of the equality sign, thus starting off with

$$\left.\begin{array}{l} .565 \cdot 10^6 x + .345 \cdot 10^6 y + .222 \cdot 10^3 z = 1,\ 0,\ 0 \\ .345 \cdot 10^6 x + 1.190 \cdot 10^6 y - .492 \cdot 10^3 z = 0,\ 1,\ 0 \\ .222 \cdot 10^6 x - .492 \cdot 10^6 y + 2.500 \cdot 10^3 z = 0,\ 0,\ 1 \end{array}\right\} \qquad (73)$$

The letters x, y, z simply designate the three unknowns that are to be solved for. Since there are three constant columns on the right, there are three different solutions. The simplest way to obtain them would be to follow the regular routine for solving normal equations illustrated in section 12b and written out in symbols in section 17. By whatever method carried out, the results written down in the order corresponding to Eqs. 73 give the reciprocal matrix

$$A^{-1} = \left|\begin{array}{ccc} 2.458 \cdot 10^6 & -0.874 \cdot 10^6 & -0.390 \cdot 10^3 \\ -.875 \cdot 10^6 & 1.226 \cdot 10^6 & .319 \cdot 10^3 \\ -.391 \cdot 10^3 & .319 \cdot 10^3 & .498 \end{array}\right| \qquad (74)$$

The occasional failure of symmetry in the 3d decimal places comes from not carrying more figures; but what we have is good enough. Supposing that the Lagrange multipliers have not been worked out, we should next compute them from Eqs. 71 as follows:

$$\left.\begin{array}{rl} \lambda_1 & = -0.133\cdot 2.458 + 0.054\cdot 0.875 - 0.073\cdot 0.391 = -0.309 \\ \lambda_2 & = + \;\text{"}\;\; 0.874 - \;\text{"}\;\; 1.226 + \;\text{"}\;\; 0.319 = 0.073 \\ 10^3\lambda_3 & = + \;\text{"}\;\; 0.390 - \;\text{"}\;\; 0.319 + \;\text{"}\;\; 0.498 = 0.071 \end{array}\right\} \quad (75)$$

These agree well enough with the values -.308, .073, and .071 already found in section 12b (conclusion 2°, p. 67).

The chief aim at present is to compute the auxiliary constants B', B", B‴ for each of the three G functions of section 12b. Going back to table 3 in section 12b for the coefficients needed for Eqs. 72 we find that

For G^1

$$\left.\begin{array}{rl} B' & = 10^3\{0.182\cdot 2.458 - 0.182\cdot 0.875 - 0.500\cdot 0.391\} = 0.0933\cdot 10^3 \\ B'' & = 10^3\{- \;\text{"}\;\; 0.874 + \;\text{"}\;\; 1.226 + \;\text{"}\;\; 0.319\} = 0.224\cdot 10^3 \\ B''' & = \quad - \;\text{"}\;\; 0.390 + \;\text{"}\;\; 0.319 + \;\text{"}\;\; 0.498 = 0.236 \end{array}\right\} \quad (76)$$

These values used in Eq. 69 give

$$\begin{aligned} 1/\text{wt. of } G^1 &= 0.500 - 0.182\cdot 0.0933 - 0.182\cdot 0.224 - 0.500\cdot 0.236 \\ &= 0.324 \end{aligned} \qquad (77)$$

That is, the weight of $G^1 = 1/0.324$, in agreement with conclusion 4° in section 12c, page 68.

For G^2

$$\left.\begin{array}{lllllll} B' = 10^3\{ & 0.222\cdot 2.458 & + & 0.492\cdot 0.875 & - & 2.500\cdot 0.391\} & = 0.00133\cdot 10^3 \\ B'' = 10^3\{- & \text{"} \quad 0.874 & - & \text{"} \quad 1.226 & + & \text{"} \quad 0.319\} & = 0.00028\cdot 10^3 \\ B''' = \quad - & \text{"} \quad 0.390 & - & \text{"} \quad 0.319 & + & \text{"} \quad 0.498 & = 1.0015 \end{array}\right\} \quad (78)$$

These used in Eq. 69 give

$$1/\text{wt. of } G^2 = 2.500 - 0.222\cdot 0.00133 - 0.492\cdot 0.00028 - 2.500\cdot 1.0015$$
$$= -0.004 \quad (79)$$

Since weights can not be negative, we may suppose that this negative result is accidental from not carrying enough figures. The lowest possible result, if all figures had been carried, would be 0. Since we know this is actually what it ought to be, we shall not dwell further on it, but shall call the result 0, whereupon the weight of G^2 is infinity, as is already known.

For G^3 As an exercise, the student should calculate B', B", and B''' for G^3 in like manner, obtaining

$$\left.\begin{array}{l} B' = -1.808 \cdot 10^9 \\ B'' = 0.865 \cdot 10^9 \\ B''' = 0.293 \cdot 10^6 \end{array}\right\} \quad (80)$$

whereupon

$$1/\text{wt. of } G^3 = 10^{12}\{3.687 + 0.657\cdot 1.808 + 0.262\cdot 0.865 - 0.094\cdot 0.293\}$$
$$= 2.300\cdot 10^{12} \quad (81)$$

in agreement with conclusion 4° in section 12c (top of page 69).

Remark The number of auxiliary constants B', B", B"', etc. in Eq. 69 is equal to the number ν of conditions, i.e. the number of F functions. This is also the number of Lagrange multipliers (λ), the number of equations in Eqs. 73, and the order of the reciprocal matrix. The number of G functions whose weights are wanted may be any whatever, smaller or larger than the number of F functions.

III. APPLICATION TO CURVE FITTING

13. *Adjustable parameters enter the conditions; the problem of curve fitting*. The problem is as heretofore, to minimize χ^2. The solution is already contained in Eqs. 59-63, pp. 36-38. The L factors now will be found simpler, consisting of fewer terms (compare Eqs. 60 and 85); but parameters may enter these terms and appear otherwise in the normal equations. There will be a function to be fitted--a relation between x and y involving unknown parameters; however expressed we transpose it all to one side of the equation and obtain the curve

$$F(x, y; a, b, c) = 0 \tag{82}$$

This will be the calculated curve in Figs. 8 and 9, pp. 82 and 83. a, b, and c are the least squares estimates of the true unknown parameters α, β, and γ. It is assumed that if there were no errors of observation, the observed coordinates X_h, Y_h of Fig. 9 would satisfy $F(x, y; \alpha, \beta, \gamma) = 0$ exactly. In other words, we proceed as if we are fitting the right curve.

For simplicity and definiteness, the development will be written out for two coordinates, x and y, at each point. The extension to three coordinates is obvious, in which event Eq. 82, instead of being the equation of a curve in the x, y plane, is written as the surface $F(x, y, z; \alpha, \beta, \gamma) = 0$ in the x, y, z space. See example 3 of section 20 for an illustration in three dimensions, and exercise 24 of section 19 for one in four dimensions.

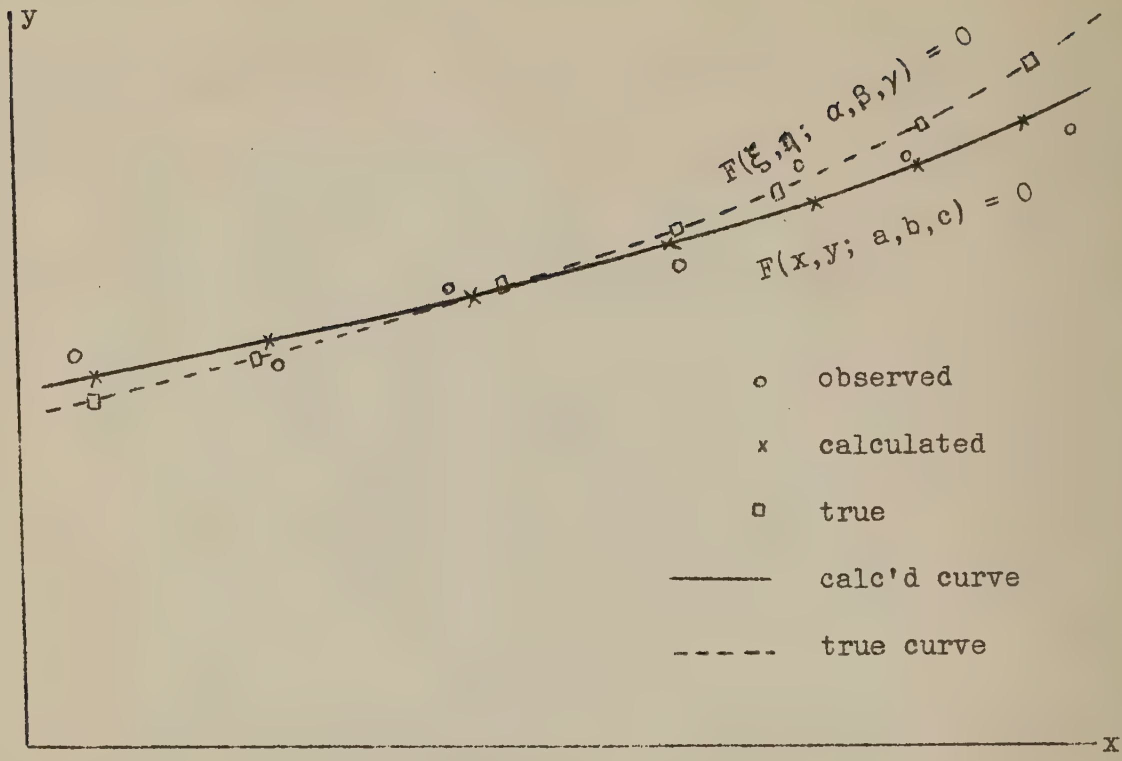

Fig. 8. A Typical Situation in Curve Fitting

It is assumed that the "true points" wherever they are lie on the "true curve" $F(\xi, \eta; \alpha, \beta, \gamma) = 0$; α, β, γ being the true and unknown values of the parameters. The "calculated points" all lie on the "calculated curve" $F(x, y; a, b, c) = 0$; a, b, c being the calculated (estimated) values of the parameters. The calculated points are estimates of the true points.

(This figure and the next one appeared in an article by the author entitled "On the chi test and curve fitting", in the J. Amer. Stat. Assoc. 29, 372-382, 1934)

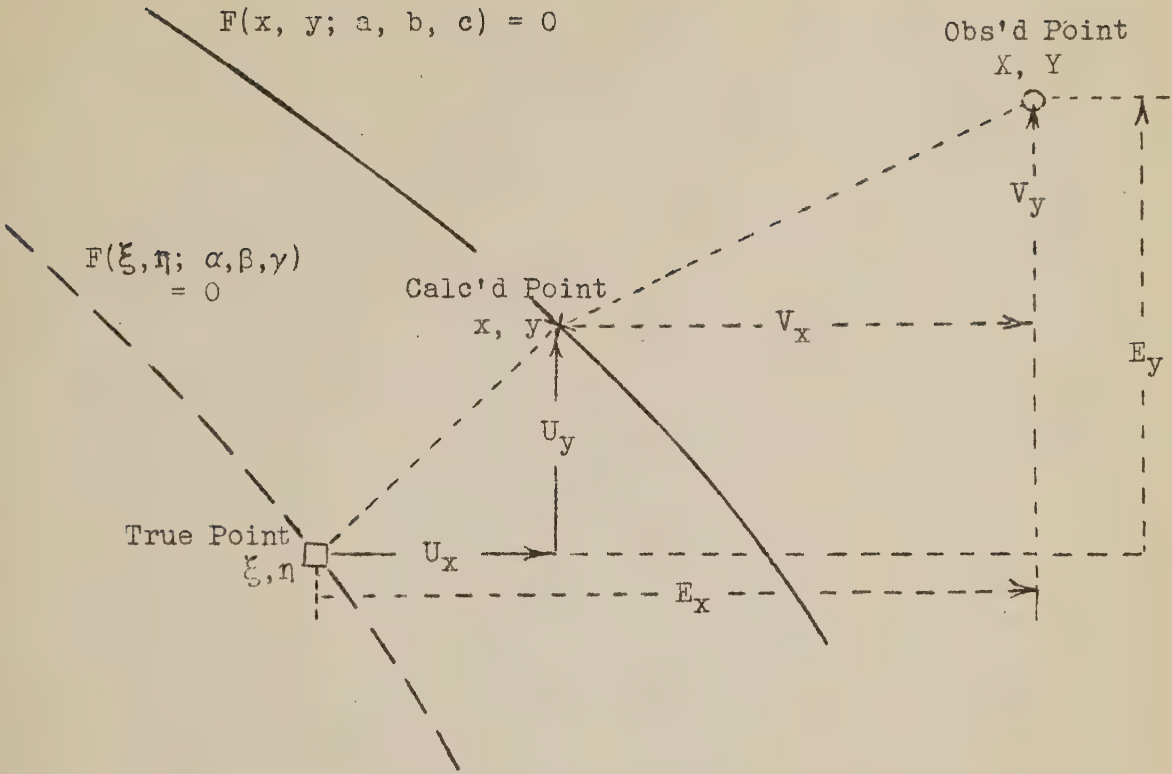

Fig. 9. Relations between the "True", "Observed" and "Calculated" Points

The x and y coordinates of a point are observed; these observations when plotted give the "observed point". The point that was measured is the "true point", which is unknown and lies on the unknown "true curve" $F(\xi, \eta; \alpha, \beta, \gamma) = 0$. α, β, γ being the true but unknown values of the parameters. The "calculated curve" is found by adjusting a series of observed points; its equation will have the same form as the true curve, but the parameters therein will be the "calculated parameters" a, b, c. Corresponding to each observed point there will be a "calculated point", whose coordinates are found by subtracting the "residuals" V_x and V_y from the observed coordinates X and Y. E_x and E_y denote the "errors in the observed point"; E_x, E_y, U_x, and U_y are unknown, but V_x and V_y are calculated along with the parameters a, b, c by the method of least squares. As the figure happens to be drawn, each of the six quantities E_x, E_y, U_x, U_y, V_x, V_y is positive. Their signs are indicated by the directions of the arrows.

A point is observed to be X_h, Y_h, i.e. the x coordinate of some point ξ_h, η_h is measured, perhaps several times, and the mean of these measurements is X_h with weight w_{xh}. Likewise, the y coordinate of the same (true) point is measured, perhaps several times, and the mean of them is Y_h with weight w_{yh}. By Eq. 16, p. 13, $w_{xh} : w_{yh} = \text{Var. } Y_h : \text{Var. } X_h$.

In order to make use of the solution already worked out, and to conform to the notation used in the preceding pages, Y_h will be written X_{n+h}, w_{xh} as w_h, and w_{yh} as w_{n+h}, there being n observed points altogether.

The conditions (50) will be n in number; i.e. the number v used in Art. 9 is now equal to n, the number of observed points. The reason is that for each observed point, there is a calculated point, and the calculated point is forced to lie <u>on the calculated curve</u>; i.e. the residuals must be just the distances required to put the calculated point on the calculated curve; see Figs. 8 and 9, also Fig. 13, p.168. If x_h, x_{n+h} (i.e. x_h, y_h) is the <u>calculated</u> point corresponding to the <u>observed</u> point X_h, X_{n+h} (i.e. X_h, Y_h), then x_h is the least squares estimate of ξ_h, and x_{n+h} (or y_h) is the least squares estimate of η_h.

Now since the calculated point x_h, x_{n+h} must lie on the calculated curve (82), then

$$F(x_h,\ x_{n+h}\,;\ a,\ b,\ c)$$

must vanish at every one of the n calculated points. Thus, for every point there is one condition imposed on the 2n adjusted coordinates x_1, x_{n+1}; x_2, x_{n+2}; ...; x_n, x_{2n}; altogether, then, there are as many conditions as there are points. For n points there are n conditions. So we set the F^h of Eqs. 50 identical with $F(x_h, x_{n+h}; a, b, c)$, i.e.

$$F^h \equiv F(x_h, x_{n+h}; a, b, c) \quad \text{or} \quad F(x_h, y_h; a, b, c) \tag{83}$$

Immediately we see that* since neither x_g nor y_g is involved at the point x_h, y_h,

$$F_g^h = 0 \quad \text{unless } g = h \quad \text{or} \quad n+h \tag{84}$$

It follows* from Eq. 60 p. 37 that $L_{gh} = 0$ unless $g = h$, whereupon the L terms off the diagonal in Eqs. 61 disappear (see Eqs. 86 ahead). For those on the diagonal Eq. 60 gives

$$L_{hh} = \frac{F_h^h \, F_h^h}{w_h} + \frac{F_{n+h}^h \, F_{n+h}^h}{w_{n+h}}$$

which is needlessly cumbrous in appearance. In the first place, since L terms will appear only on the diagonal, the double suffix is not required, so hereafter we shall write simply L_h; in fact we shall soon form the habit of omitting the suffix altogether, writing

$$L = \frac{F_x F_x}{w_x} + \frac{F_y F_y}{w_y} \tag{85}$$

it being understood that all quantities involved may vary from point to point. The first term drops out if X is free of error, the second if Y is free of error; see Remark 2 in Ex. 4 of section 19, p. 125.

By Eq. 48 p. 27 we see that L at the point h is actually the weight of $F(X_h, Y_h; a, b, c)$ or of the quantity called S in exercises 3, 4, and 5 at the end of **section 15**. On this account, W will

---

* I have seen two problems in curve fitting wherein this was not true; the coordinates of one point depended by addition or multiplication on the coordinates of an adjacent point. One of these is worked out in the Phil. Mag. (London) 17, 804-829, 1934.

frequently be written in place of $1/L$ as we go along, though it is better to perform the actual numerical calculations with L rather than with its reciprocal (cf. section 16 and some remarks in exercise 4 of section 19).

The general normal equations (p. 37) for curve fitting now take the form shown below in Eqs. 86. They can quickly be reduced to a smaller set, Eqs. 88, in number equal to the number of adjustable parameters.

λ_1	λ_2	λ_3	$\cdots$	λ_n	v_a	v_b	v_c	$= 1$
L_1	0	0	$\cdots$	0	F_a^1	F_b^1	F_c^1	F_o^1
0	L_2	0	$\cdots$	0	F_a^2	F_b^2	F_c^2	F_o^2
0	0	L_3	$\cdots$	0	F_a^3	F_b^3	F_c^3	F_o^3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0	0	0	$\cdots$	L_n	F_a^n	F_b^n	F_c^n	F_o^n
F_a^1	F_a^2	F_a^3	$\cdots$	F_a^n	0	0	0	0
F_b^1	F_b^2	F_b^3	$\cdots$	F_b^n	0	0	0	0
F_c^1	F_c^2	F_c^3	$\cdots$	F_c^n	0	0	0	0

(86)

$\lambda_1, \lambda_2, \ldots, \lambda_n$ are easily eliminated; solve for them in the upper n equations, getting

$$\left.\begin{aligned} \lambda_1 &= \frac{1}{L_1}(F_o^1 - F_a^1 v_a - F_b^1 v_b - F_c^1 v_c) \\ \lambda_2 &= \frac{1}{L_2}(F_o^2 - F_a^2 v_a - F_b^2 v_b - F_c^2 v_c) \\ &\vdots \quad \vdots \\ \lambda_n &= \frac{1}{L_n}(F_o^n - F_a^n v_a - F_b^n v_b - F_c^n v_c) \end{aligned}\right\} \quad \text{(A } \lambda \text{ for each point)} \qquad (87)$$

and then substitute these values of $\lambda_1, \lambda_2, \ldots, \lambda_n$ into the lower three equations of (86). The result is

$$\begin{array}{cccccc}
v_a & v_b & v_c & = & 1 & \\
\hline
\left[\frac{F_aF_a}{L}\right] & \left[\frac{F_aF_b}{L}\right] & \left[\frac{F_aF_c}{L}\right] & & \left[\frac{F_aF_o}{L}\right] & \\
\left[\frac{F_aF_b}{L}\right] & \left[\frac{F_bF_b}{L}\right] & \left[\frac{F_bF_c}{L}\right] & & \left[\frac{F_bF_o}{L}\right] & \qquad (88) \\
\left[\frac{F_aF_c}{L}\right] & \left[\frac{F_bF_c}{L}\right] & \left[\frac{F_cF_c}{L}\right] & & \left[\frac{F_cF_o}{L}\right] &
\end{array}$$

These equations contain only the residuals v_a, v_b, v_c as unknowns. The arrangement of the coefficients is symmetrical and their quadratic form positive definite, like the general normal equations whence they came (section 9). Once the residuals v_a, v_b, and v_c are obtained, the adjusted values of the parameters are found immediately by subtraction from the approximate values, i. e.

$$\begin{aligned} a &= a_o - v_a \\ b &= b_o - v_b \qquad\qquad [\text{Eqs. 62, p.38}] \\ c &= c_o - v_c \end{aligned}$$

As noted in section 9, the final values of a, b, and c will be the same, regardless of the approximations a_o, b_o, and c_o. In some problems, however, these approximations must not be too rough; in others, it makes no difference what they are, except that more figures are required in the normal equations; see exercises 4, 5, and 10 in section 19. In regard to the matter of arriving at the approximations a_o, b_o, c_o, see a footnote in section 9 on the "method of selected points", also example 2 in section 20.

Thus the calculated curve (Fig. 9) is fixed; it is given by Eq. 82 with the values of a, b, and c just written.

Several details remain: to adjust the observations (sections 14 and 15); to work out a systematic procedure for setting up the normal equations and solving them (sections 16 and 17); to calculate the weights, or the variance and product variance coefficients of a, b, c (the reciprocal matrix, sections 17 and 18); a short-hand calculation of the minimized value of ϕ^2 (section 17).

Remark It is well known that a rapid method of computing the mean $\bar{x}$ and the S. D. s of a set of n observations $x_1, x_2, \dots, x_n$, is to select first of all an arbitrary datum, perhaps a rounded off guess at $\bar{x}$, then to write down the departures from this datum, and their squares, taking finally their averages, and correcting the arbitrary datum to find $\bar{x}$, and at the same time s. This is exactly equivalent to fitting the line

$$x = a \qquad \text{(see section 3)}$$

by least squares, the arbitrary datum a_o being considered an approximation for a. The work might be laid out something like this

Observations (notation of section 3)	Departure from a_o	Square of the departure from a_o
x_1	$x_1 - a_o$	$(x_1 - a_o)^2$
x_2	$x_2 - a_o$	$(x_2 - a_o)^2$
x_3	$x_3 - a_o$	$(x_3 - a_o)^2$
$\vdots$	$\vdots$	$\vdots$
x_n	$x_n - a_o$	$(x_n - a_o)^2$
	Av. = A (say)	Av. = B (say)

$\bar{x}$ and s are calculated by taking

$$\bar{x} = a_o + A$$

$$s^2 = B - A^2$$

The student should prove that these equations are true to the definitions of $\bar{x}$ and s, no matter what value be taken for a_o. The process is easily modified to take care of unequal numbers of observations (unequal weights).

Numerical examples of this procedure abound. R. A. Fisher's Statistical Methods for Research Workers contains many applications, and it is common in most all papers dealing with the analysis of variance, as the student is doubtless aware. An example is worked out in Whittaker and Robinson's Calculus of Observations (Blackie and Son, 1929) Art. 97; one also in Deming and Birge's Statistical Theory of Errors. The reason for mentioning it here is merely to emphasize that fundamentally the procedure amounts to a least squares solution wherein a_o is an approximate value of a (actually a_o need not be anywhere near the final value of a, but the computation is easier if is).

After sections 16 and 17 have been studied, it will be interesting to return to the following tabular scheme, which is equivalent to the procedure just outlined for the computation of the mean and S. D. of a set of n observations. We form the sums $\Sigma(x - a_o)$ and $\Sigma(x - a_o)^2$ calling them nA and nB respectively, as above, and enter them in Eqs. I and 2 below, performing on them the steps indicated.

No.	v_a	=	1	C	
I	n		$-nA$	1	
2			nB	0	
3			$-nA^2$	...	(Multiply I thru by $\frac{nA}{n}$)
II			$nB - nA^2$	...	(Add 2 and 3)

Solving Eq. I for v_a we get $v_a = -A$, whence the least squares value of a is

$$a = a_o - v_a = a_o + A$$

$$= a_o + (\bar{x} - a_o) = \bar{x}$$

Looking at the left-most entry in II we see $nB - nA^2$, which is none other than ns^2. The tabulation seen here is a simple extension of the one given in section 4b, and a rather trivial application of the routine process to be learned in section 17 for the solution of normal equations and the computation of the sum of squared residuals in curve fitting.

14. Adjusting the observations, or finding the calculated points. Going back to Eqs. 58, p. 36 we see that the residuals will be

$$\left.\begin{aligned} V_x &= \frac{1}{w_x}\lambda_h F_x \qquad \text{(x residual at point h)} \\ V_y &= \frac{1}{w_y}\lambda_h F_y \qquad \text{(y residual at point h)} \end{aligned}\right\} \tag{89}$$

The adjusted (calculated) coordinates can now be found; they are

$$\left.\begin{aligned} x_h &= X_h - V_x \\ y_h &= Y_h - V_y \end{aligned}\right\} \quad \text{at point h (see bottom p. 38)} \tag{90}$$

These are the coordinates of the calculated point corresponding to the observed point X_h, Y_h. Finding the calculated points is the process of adjusting the observations; when x_h and y_h have been calculated, the observations X_h, Y_h are said to be adjusted. The calculated point x_h, y_h is the least squares estimate of the position of the unknown true point ξ_h, η_h. Obviously, it will depend not only upon X_h, Y_h and their weights, but also more or less upon all the other points and their weights.

Just how is this dependence tied up with the other points? Through the normal Eqs. 86, or their equivalent, Eqs. 87 and 88: Eqs. 88 supply the parameter-residuals v_a, v_b, and v_c, which are to be used in Eqs. 87 to find $\lambda_1, \lambda_2, \ldots, \lambda_n$. These in turn are used in Eqs. 89 to compute the x and y residuals at each point, by which the observations are adjusted as indicated in Eq. 90.

We now have a method of adjusting the observations and of estimating the parameters a, b, c, when both the x and y coordinates are subject to error, but it must be remembered that the solution depends on certain simplifying assumptions; see footnote on pp. 34-35, also the exercises of section 19.

A valid and familiar method of adjusting the observations, available only when all the x coordinates or all the y coordinates are free of error, is to substitute the coordinate free of error into the formula $F(x_h, y_h; a, b, c) = 0$, and solve for the other coordinate. Thus, in the parabola $y = a + bx + cx^2$, if x is free of error it is easy to calculate y for a given x, once a, b, and c are determined. But if only y is subject to error, only V_y and y_h need to be calculated from Eqs. 89 and 90. Likewise, if only x is subject to error, only V_x and x_h need to be calculated. When the function F of Eq. 82 is somewhat involved, even though only one coordinate is subject to error it will sometimes be easier to adjust the observations, i.e. to compute the "calculated points", by means of the Lagrange multipliers, through Eqs. 87, 89, and 90, than to substitute directly into the formula $F(x_h, y_h; a, b, c) = 0$ and solve it for the desired coordinate in terms of the one that is given. When both coordinates are subject to error, one must apply Eqs. 87, 89, and 90, if he would adjust the observations. For a numerical illustration, see pages 165, 166, and 168: it is suggested there that the student work out the numerical values of the remaining nine calculated points.

It might be well to look for a moment at the parabola $y = a + bx + cx^2$ again, and consider the difficulty of solving for x in terms of y, as one would need to do by the older method if only x were subject to error. Of course it can be done, but it would be easier to compute V_x at each point by means of Eqs. 87, 89, and 90.

In the past, a great deal of emphasis has been placed on the parameters. Books and courses on <u>adjustment of observations</u> have usually dealt with estimating the parameters a, b, c, hardly mentioning the adjustment of observations. Least squares is primarily a method of adjusting the observations by minimizing χ^2. The parameters enter only as unknowns in the conditions that are forced upon the adjusted observations. As a matter of fact, <u>least squares is the only method of curve fitting that professes to adjust the observations</u>.

Now of course, the estimates of the parameters and their S. Es. are often the prime purpose of an investigation. But however important they seem to be, it is well to keep in mind just how and why they enter the adjustment by least squares, in order to understand the results.

This brings up another point of interest. The estimate of σ by external consistency; called σ(ext) heretofore (section 6c) depends upon the calculation of ϕ^2 or $\Sigma(w_x V_x^2 + w_y V_y^2)$. Neither the fit of the curve, nor the estimated S. Es. of the parameters, both depending on ϕ^2, can be calculated unless both the x and y coordinates are given their correct relative weights.

It is a fact that no matter how many coordinates at any point are subject to error, the least squares value of

$$\chi^2 = \frac{1}{\sigma^2}\Sigma(w_xV_x^2 + w_yV_y^2) \tag{91}$$

for the fitted curve has the probability distribution

$$p(\chi^2)\ d\chi^2 = \frac{1}{\Gamma(\frac{1}{2}k)\,2^{\frac{1}{2}k}}(\chi^2)^{\frac{k-2}{2}}\ e^{-\frac{1}{2}\chi^2}\ d\chi^2 \tag{92}$$

with k = number of points - number of adjustable parameters. The proof of the χ^2 distribution in the general case is contained in the paper cited in the legend of Fig. 8. A necessary lemma thereto is given in the London Phil. Mag. 19, 389-402, 1935. The effect of including both x and y residuals in χ^2 is merely to increase the value of χ^2 over what it would be if only one coordinate were subject to error, the form of the curve being unaffected.

It is to be noted that there is no such thing as a distribution of χ^2 unless the fitting is done by least squares; in other words, only the minimized χ^2 has a distribution. (The assumption of normally distributed observations is basic to all these statements).

The adjustment of the observations, or what amounts to the same thing, the calculation of V_x and V_y at all points, is important; to know the sum of $w_xV_x^2 + w_yV_y^2$ is often not enough and it is then highly desirable to make a study of the individual deviations. See the discussion in Example 2 of section 20.

15. Some geometry concerning the adjustment of observations.

Now let us consider some of the details connected with the calculated points, or the adjusted observations. The accompanying diagram and analysis are helpful to an understanding of just what is accomplished by the application of least squares.

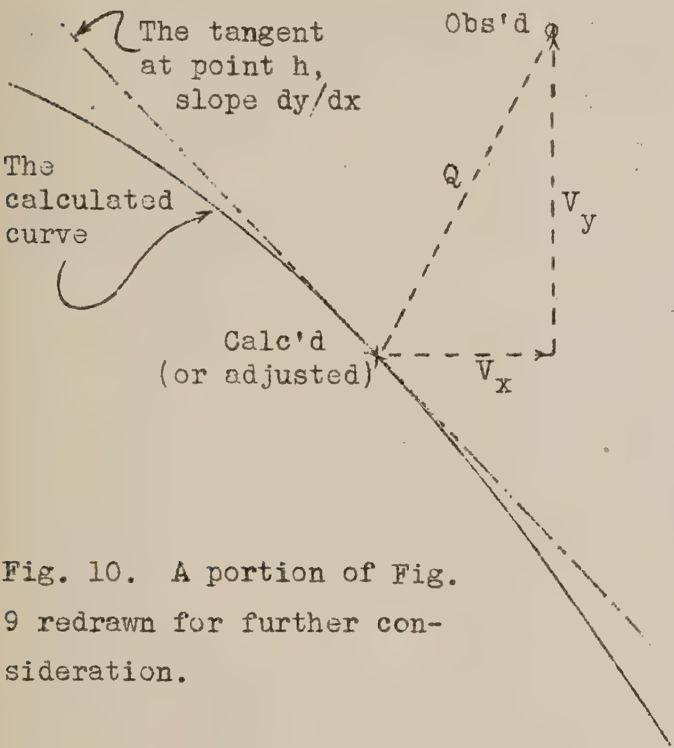

Fig. 10. A portion of Fig. 9 redrawn for further consideration.

Let Q be the line segment joining the observed and calculated points. By Eqs. 89 we can find the slope of this line segment; it is

$$\text{Slope of Q} = \frac{\text{the y residual}}{\text{the x residual}} = \frac{V_y}{V_x} = \frac{w_x}{w_y}\,\frac{F_y}{F_x} = \frac{w_x}{w_y}\,\frac{\dfrac{dF}{dy}}{\dfrac{dF}{dx}} = -\,\frac{w_x}{w_y}\,\frac{dx}{dy} \qquad (93)^*$$

That is,

$$\text{The slope of Q} = -(w_x/w_y)\,\frac{1}{\text{the slope of the curve at h}} \qquad (94)$$

---

* The last step here involves the very important relation learned in elementary calculus, that if $F(x,y) = c$, then $dy/dx = -\dfrac{dF}{dx}\Big/\dfrac{dF}{dy} = -F_x/F_y$.

Hence if at point h, $w_x = w_y$, the two slopes are negative reciprocals of one another and <u>the line segment Q is perpendicular to the curve</u> (but see exercises at the end of this section).

If at any point, $w_x : w_y = \infty$ or is practically very large, which is to say that X is relatively infallible, then the line segment Q is vertical and the adjustment is all in the y coordinate.

If at any point, $w_y : w_x = \infty$ or is practically very large, which is to say that Y is relatively infallible, then the line segment Q is horizontal and the adjustment is all in the x coordinate.

If at any point, $w_x : w_y$ is finite, i.e. both X and Y are subject to error, the line segment Q will be neither horizontal nor vertical, but inclined, and there is adjustment in both the x and y coordinates.

<u>Exercise 1</u>. With x and y measured in certain units, the weights of X and Y at a certain point are equal, and the line segment Q is therefore perpendicular to the curve at that point. Then the units in which y is measured are changed, as from feet to inches. Prove that the line segment Q is no longer perpendicular to the fitted curve. (Hint: the ratio $w_y : w_x$ was unity before the change in scale, but it is not so afterward. If all the y coordinates are multiplied by C, then the weights of all y observations are decreased by the factor $1/C^2$, and the new value of $w_y : w_x$ is $1/C^2$ times the old; hence is no longer unity, C being supposed not unity).

Exercise 2. When there are three coordinates, the surface

$$F(x, y, z; a, b, c) = 0$$

is to be fitted to the n observed points. L then contains three terms--the two already written in Eq. 85 plus $F_z F_z / w_z$. Show that if x, y, z are observed with equal weight at any point, the line segment Q joining the observed and calculated points is normal to the fitted surface.* In such a problem, the calculated points lie on the calculated (fitted) surface. See the last part of section 20 for an example in three dimensions, and Ex. 24 of section 19 for one in four dimensions.

Exercise 3. Let S be the value that F_o would have at point h if a_o, b_o, c_o were taken equal respectively to the final values a, b, c; i.e., in symbols, let

$$\begin{aligned} S &= F_o - F_a v_a - F_b v_b - F_c v_c \\ &= F_x V_x + F_y V_y \qquad \text{(see Eq. 53, p. 35)} \\ &= F(X_h, Y_h; a, b, c) \end{aligned}$$

Then $\Sigma(w_x V_x^2 + w_y V_y^2)$ can be written $\Sigma W S^2$, and this would be the value of ϕ^2 calculated under the assumption that $a = a_o$, $b = b_o$, $c = c_o$. W is here written for 1/L in Eq. 85. Hint: From Eqs. 87, $\lambda = WS$, whence Eqs. 89 give

$$V_x = (1/w_x)\ W\ S\ F_x$$

$$V_y = (1/w_y)\ W\ S\ F_y$$

---

* This result was first stated in the Phil. Mag. (London) 11, 146-155, 1931.

for the x and y residuals at point h. Substitution into $w_x V_x^2 + w_y V_y^2$ gives the required result in terms of S (due to Kummell, 1879). This result is useful in Exercise 3 of section 17.

<u>Exercise 4</u>. Prove that W in the preceding exercise is actually the weight of S, i. e. of $F(X_h, Y_h; a, b, c)$. (Hint: Apply Eq. 48, p. 27). Hence the new expression $\Sigma\, W\, S^2$ for ϕ^2 can be regarded as a sum of the weighted squares of residuals, S now being defined as a new kind of residual.

<u>Exercise 5</u>. If the x coordinate is free of error, WS^2 is equal to $w_y V_y^2$, and if y is free of error, WS^2 is equal to $w_x V_x^2$.

<u>Exercise 6</u>. Prove that for any observed point in the neighborhood of which the slope of the fitted curve is positive, the residuals V_x and V_y will have opposite signs; but if the slope is negative, then V_x and V_y will have the same sign. In other words, when the fitted curve lies below the observed point, then the calculated point lies below and to the right of the observed point if the slope of the fitted curve is positive, but below and to the left if the slope is negative; and when the fitted curve lies above the observed point, then the calculated point lies above and to the left if the slope is positive, above and to the right if the slope is negative.

16. Systematic procedure for setting up the normal equations for the parameters. Given the formula to be fitted, one can transpose it all to one side of the equation and have

$$F(x, y; a, b, c) = 0 \tag{82}$$

The first step* is to work out somehow satisfactory approximations a_o, b_o, c_o for the parameters (cf. some remarks in section 9; also to calculate F_o at every point). In some problems, depending on the formula and the weighting, it is permissible** to take $a_o = b_o = c_o = 0$ when calculating F_o (but not L), in which event the residuals v_a, v_b, v_c turn out to be the adjusted values -a, -b, -c themselves. But this is not usually advisable even when permissible. As a matter of saving time a good rule is to commence the adjustment with as good approximations a_o, b_o, c_o as can be found with a reasonable amount of trouble, and thus to cut down the number of figures required in the formation and solution of the normal equations.

The second step requires some differential calculus; it consists merely in writing down the derivatives of F, namely

$$F_a,\ F_b,\ F_c,\ F_x,\ \text{and } F_y$$

looking toward the calculation of

$$L = F_x F_x/w_x + F_y F_y/w_y \qquad \text{(refer back to Eq. 85 p. 85)}$$

---

* It is interesting to compare these steps and tables with those of section 12a, where there were no parameters.

** See Exs. 5, 10, 17, and 21 in section 19.

and the summation required at the n points. L may vary from point to point, and some or all of the derivatives F_a, F_b, and F_c almost surely will. So will F_o.

The <u>third step</u> is to work out the numerical values of F_a, F_b, F_c, and L, at every point. The following tabulation is suggested.

Table 1 (3d step)

h	F_x	w_x	F_y	w_y	L	$\sqrt{L}$	F_a	F_b	F_c	F_o
1	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-	-	-
⋮										⋮
n	-	-	-	-	-	-	-	-	-	-

Of course auxiliary columns may be required, depending on the problem and the whims of the computer. Or, perhaps some columns listed will not be needed, e. g. if w_x were ∞ all the way down (x free from error) then F_x and w_x would be omitted, since y alone would contribute to L, which would be merely F_yF_y/w_y. Likewise, if y were free from error all the way down, then the F_y and w_y columns would not be needed, for then L would be simply F_xF_x/w_x.

The <u>fourth step</u> is to divide each entry under F_a, F_b, F_c, and F_o by the corresponding $\sqrt{L}$. The sums at the right or bottom (one but not both) of table 2 can be formed by cumulating these quotients in the horizontal or vertical, the individual quotients being entered in the table. (This cumulation requires a machine with a double multiplying

dial, one to be locked for cumulating quotients, while the other clears when desired. See a remark following table 2 in section 12a, page 54.)

Table 2.--The matrix for the formation of the normal equations*
(4th step).

h	$\frac{F_a}{\sqrt{L}}$	$\frac{F_b}{\sqrt{L}}$	$\frac{F_c}{\sqrt{L}}$	$\frac{F_o}{\sqrt{L}}$	Sum
1	-	-	-	-	-
2	-	-	-	-	-
3	-	-	-	-	-
⋮					⋮
n	-	-	-	-	-
Sum	-	-	-	-	-✓

The sums at the right and along the bottom are used for checking the formations of the normal equations exactly as was done with table 2 in section 12a. First of all, the sum across the bottom should equal the sum down the right-hand side, as indicated by the check mark. In running down the columns, cumulating squares and cross-products (the 5th step, next page), the final total in the multiplier register will equal the sum at the bottom of the multiplier column provided no changes in sign occur in the multiplicand column. In a machine with a double multiplier register, one part of which can be locked for cumulation while the other clears, individual multipliers can be checked at will in one dial, while the sum of the multipliers cumulates in the other one for checking at the bottom.

---

* Concerning the use of the term matrix here, see the footnote on p. 161.

It is difficult to imagine circumstances wherein a maximum of three or four significant figures in any column will not suffice. This means that if there is great variation in the vertical in any column, some entries in table 2 may have only two or one, or not even any figures; see, for instance, pages 154 and 161. The denominations of the different columns should be made uniform by writing powers of 10 at the top of each row, to apply to the whole column (see the solved examples at the end; also the one in section 12b). No attention need be given to the powers of 10 until the end, when the solution of the normal equations is decoded (unscrambled, I have called it in my lectures).

The fifth step is to form the normal equations (see the next section) from table 2 by the familiar process of adding squares and cross-products of columns. Thus, no matter how complicated the weighting, and no matter what be the form of the fitted curve, the whole procedure is uniform, and we are brought to a uniform and familiar process for the formation of the normal equations.

As already suggested, the student should compare this matrix with the previous table 2 of section 12a, which arose in the consideration of conditions not containing parameters. The headings in the tables are different there, of course; but the process of forming the normal equations from table 2 is the same here as it was there-- in each case a routine procedure. Also, the routine of solution is the same (compare sections 12b and 17). The exercises in section 19 will provide practice in the necessary steps for setting up the normal equations for several types of functions.

Important note. By the procedure here explained for the formation of table 2, whence the normal equations are to be set up, the solution (i. e. the values of v_a, v_b, v_c, ϕ^2, etc.) is unequivocal. That is, it does not matter in what form the equation to be fitted is written. If one had, for example, $y = ae^{bx}$, he could put the same equation in the form $\ln y = \ln a + bx$ using in the former case $F = ae^{bx} - y$ and in the latter case, $f = \ln a + bx - \ln y$. These things will be made clear in the exercises of section 19. The point that I am making is that when the normal equations are correctly made up, they will yield the same results from any form of the fitted equation, to within higher powers of the residuals. As another example, consider the straight line in the two forms $y = a + bx$, and $x = -a/b + y/b$; again the results would be the same to within higher powers of the residuals. Summed up, the results--the final calculated parameters, and the adjusted observations, are independent of the form in which the equation is written.* Many textbooks have been indefinite and ambiguous in matters of this nature, though it is a pleasure to record that Leland in his Practical Least Squares (McGraw-Hill, 1921) states clearly that the weight of $\ln y$ is y^2 times the weight of y (see exercises 17, 18, and 21 in section 19).

---

* Very large residuals, i. e., very rough data, will invalidate this statement; see some other remarks in section 9, also in Exs. 18 and 22 of section 19.

17. Systematic solution of the normal equations. The reciprocal matrix. Systematic computation of ϕ^2. As has already been noted, the sums of squares and cross products occurring in the normal equations 88 (p. 87) are formed directly from table 2 of the preceding section. Thus $\left[\frac{F_aF_a}{L}\right]$ is the sum of squares under the column headed $\frac{F_a}{\sqrt{L}}$; the summation $\left[\frac{F_aF_b}{L}\right]$ is the cumulation of cross products under the columns headed $\frac{F_a}{\sqrt{L}}$ and $\frac{F_b}{\sqrt{L}}$; etc.

The following systematic solution of the normal equations is the same one that was used in section 12b, save for slight departures toward the end to accomplish different purposes; also it is the same scheme that was illustrated in the simple problems of section 4b, 7b, and at the end of section 13. The work is checked at pivotal points as it progresses (shown by ✓).

In the left-most entry of Eq. IV is found the minimized value of ϕ^2. The comparison of this value of ϕ^2 with that calculated by summing the weighted squares of the individual residuals V_x and V_y gives a check on the entire process, from table 1 clear through to the end (Whittaker and Robinson, Calculus of Observations, Art. 118). The check, however, is not very sensitive.

The reciprocal matrix is calculated in Eqs. 11, 12, and 13, in the columns headed C_1, C_2, C_3; it contains the variance and product variance coefficients for the parameters a, b, c; also is useful as a multiplier for solving the normal equations with any constant ("1")

column whatever, as has been known for many years;* see the exercises following, also the numerical illustration in section 12d. It is not generally known, though, that the "reciprocal solution" found by using the reciprocal matrix as a multiplier is a very sensitive indicator of instability, and that it therefore breaks down in the case of near indeterminacy**--a fact that detracts rather drastically from its usefulness in the solution of normal equations in curve fitting, where near indeterminacy is surprisingly common. In extreme cases it can be recognized by the "freezing" of the solution--the near vanishing of the last coefficient (the left-most entry [cc.2] in Eq. III in the case of three parameters); also by a very small value of A, the determinant of the coefficients, and the correspondingly large numbers found in the reciprocal matrix. (For an easy evaluation of the determinant of the coefficients, see exercise 1 following). Further discussion occurs in example 1 of section 20.+

The short-hand notation [aa] for $\left[\frac{F_aF_a}{L}\right]$, [ab] for $\left[\frac{F_aF_b}{L}\right]$, [oo] for $\left[\frac{F_oF_o}{L}\right]$, etc. will be introduced here for convenience in writing. The normal equations 88, p. 87, with the unit matrix in the C_1, C_2, C_3 columns, are solved according to the scheme on page 107, which comes with modifications from Leland's Practical Least Squares,

---

* The earliest reference that I know of on the use of the reciprocal matrix as a multiplier for solving a set of linear equations is Art. 8 of Gauss' Supplementum Theoriae Combinationis Erroribus Minimis Obnoxiae (Göttingen, 1826; Werke, vol. 4).

** W. Edwards Deming, Science, 7th May 1937.

\+ See also A. de Forest Palmer Theory of Measurements (McGraw-Hill, 1912) p. 77.

and is similar to Doolittle's* solution, which in turn goes back to Gauss**. The Gauss symbols [bb.1], [cc.2], etc., seen in Eqs. II and III will facilitate reference to certain entries later on, as in the exercises beginning on page 108. All entries below Eq. 4 arise by the simple operations of multiplication and addition indicated under "How obtained". The check marks show the "sum check" at the pivotal points. For numerical examples, see sections 12b and 20. Note that v_a is eliminated in Eq. II; v_a and v_b are both eliminated in Eq. III.

Eq. 11 comes by dividing III through by [cc.2] to get v_c.

Eq. 12 comes by substituting from 11 into II to get v_b.

Eq. 13 comes by substituting from 11 and 12 into I to get v_a.

The "back solution"

Other methods of solution have been contrived, but a complete summary is not called for here. Some of them are devices for calculating the reciprocal matrix to be used as a multiplier (pp. 105 and 111), for example, T. Smith's+, and a very promising scheme of matrix squaring devised by Hotelling and Girshick on the basis of a theorem regarding the characteristic equation of a determinant, and now being critically tested by Mr. Girshick in the Bureau of Home Economics. In another direction there is Kelley and Salisbury's++ ingenious acceleration of an iterative process usually known as Seidel's (1874) though described earlier by Gauss and Jacobi°, the same being particularly effective when the number of normal equations is large. Then there is a fascinating pivotal

---

* M. H. Doolittle, Coast and Geodetic Survey Report for 1878 (Washington), App. 8, pp. 115-118.

** Gauss, Supplementum Theoriae Combinationis (cited on page 36) Art. 13.

\+ T. Smith, "The calculation of determinants and their minors", Phil. Mag. (London) 3, 1007-9, 1927.

++ Truman L. Kelley and Frank S. Salisbury, J. Amer. Stat. Assoc. 21, 282-292, 1926.

° Whittaker and Robinson, Calculus of Observations (Blackie & Son, 1924) Art. 130. See also pp. 30-32 of Harold Hotelling's "Analysis of a complex of statistical variables into principle components" (Warick &

(Cont'd on p. 108)

The normal equations and their solution

No.	How obtained	v_a	v_b	v_c	= 1	C_1	C_2	C_3	Sum
I		[aa]	[ab]	[ac]	[ao]	1	0	0	...✓
2			[bb]	[bc]	[bo]	0	1	0	...✓
3				[cc]	[co]	0	0	1	...✓
4					[oo]	0	0	0	...✓
5	Ix -[ab]/[aa]		$-\frac{[ab]^2}{[aa]}$	$-\frac{[ac][ab]}{[aa]}$	$-\frac{[ao][ab]}{[aa]}$	$-\frac{[ab]}{[aa]}$	0	0	...
II	2 + 5		[bb.1]	[bc.1]	[bo.1]	$-\frac{[ab]}{[aa]}$	1	0	...✓
6	Ix-[ac]/[aa]			...	...	...	...	...	...
7	IIx-[bc.1]/[bb.1]			...	...	...	...	0	...
III	3 + 6 + 7			[cc.2]	[co.2]	...	...	1	...✓
8	Ix-[ao]/[aa]				$-\frac{[ao]^2}{[aa]}$	0	0	0	...
9	IIx-[bo.1]/[bb.1]				$-\frac{[bo.1]^2}{[bb.1]}$	...	...	0	...
10	IIIx -[bo.2]/[bc.2]				$-\frac{[co.2]^2}{[cc.2]}$	...	...	...	...
IV	4 + 8 + 9 + 10				ϕ^2	...	...	...	...✓
13	I solved for v_a				v_a	c_{11}	c_{12}	c_{13}	
12	II solved for v_b				v_b	c_{21}	c_{22}	c_{23}	
11	III solved for v_c				v_c	c_{31}	c_{32}	c_{33}	...✓

process invented by A. C. Aitken* of Edinburgh in 1932, after T. Smith's method; he has now, however, superseded this solution by the introduction of a number of important refinements, as I have had the privilege of seeing. Mr. Girshick is testing also a new scheme originated by Professor Hotelling for accelerating the convergence of his iterative process described in 1933 (reference beginning at the bottom of page 106). It is also interesting to note that electrical circuit machines, capable of solving something like 10 linear equations, practically instantaneously after plugging the coefficients, are in operation and undergoing further development at several centres.

Exercise 1. (a) The determinant of the coefficients of the normal equations on page 107 can be evaluated as follows:

$$A = \begin{vmatrix} [aa] & [ab] & [ac] \\ [ab] & [bb] & [bc] \\ [ac] & [bc] & [cc] \end{vmatrix} = [aa] \cdot [bb.1] \cdot [cc.2] \qquad (95)$$

which is to say that the determinant A is the product of the left-most numbers in Eqs. I, II, and III. This is important in considerations of near-indeterminacy.

(b) Show that none of the left-most entries in Eqs. I, II, or III can be negative.

Exercise 2. The matrix reciprocal to A can be denoted by

$$A^{-1} = \begin{vmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{vmatrix} \qquad (96)$$

(a) Show that the solution of the normal equations with the constant columns C_1, C_2, and C_3 leads to the values

---

York, Baltimore, 1933; originally published in the J. Educational Psychology, Sept. and Oct. 1933).

* A. C. Aitken, "On the evaluation of determinants, the formation of their adjugates, ...," Proc. Edinburgh Math. Soc. 3, 207-219, 1932.

$$c_{11} = \frac{\text{cofactor of [aa]}}{A}, \quad c_{12} = \frac{\text{cofactor of [ab]}}{A}, \quad c_{13} = \frac{\text{cofactor of [ac]}}{A}$$

$$c_{21} = \frac{\text{cofactor of [ab]}}{A}, \quad c_{22} = \frac{\text{cofactor of [bb]}}{A}, \quad c_{23} = \frac{\text{cofactor of [bc]}}{A}$$

$$c_{31} = \frac{\text{cofactor of [ac]}}{A}, \quad c_{32} = \frac{\text{cofactor of [bc]}}{A}, \quad c_{33} = \frac{\text{cofactor of [cc]}}{A}$$

(Since A falls in the denominator of each element, a small value of A, near-indeterminacy, results in high S. Es. of the parameters; see exercises 2a and 11b of section 19)

(b) Like the determinant A of the coefficients, the reciprocal matrix A^{-1} is symmetrical, i. e. $c_{12} = c_{21}$, $c_{13} = c_{31}$, $c_{23} = c_{32}$.

(c) The matrices A and A^{-1} are also alike in another respect--the terms on the main diagonal will always be positive.

(d) Show that $c_{33} = 1/[cc.2]$ = the reciprocal of the coefficient of the 3d unknown in III.

<u>Exercise 3</u>. (a) Combine Eqs. 65 and 87 to get

$$\phi^2 = [oo] - [ao]v_a - [bo]v_b - [co]v_c$$

(b) Prove that the left-most entry in Eq. IV of the solution exhibited above is actually ϕ^2.

(c) Show also, by noting how Eqs. 8, 9, and 10 are formed, that

$$\phi^2 = [oo] - [aa]v_a''^2 - [bb.1]v_b'^2 - [cc.2]v_c^2$$

where $v_a'' = [ao]/[aa]$ = the value of v_a that would be obtained if b and c were fixed (not adjustable) at the values b_o and c_o, and wherein also $v_b' = [bo.1]/[bb.1]$ = the value of v_b that would be obtained if c were fixed at the value c_o, but a and b both adjustable.

Remark 1. This result sheds a singular elegance on the form of solution exhibited on page 107. The term [oo] seen in Eq. 4 is the sum of the weighted squares of the residuals calculated under the assumption that $a = a_o$, $b = b_o$, $c = c_o$ (see exercise 3 of section 15, p. 97). The three negative terms in Eqs. 8, 9, and 10 on page 107 are precisely the amounts subtracted from [oo] at the bottom of the preceding page, and in the same order. That is to say, by the routine solution outlined on page 107 there will appear (1°) in Eq. 8 the reduction in weighted squares that is brought about by allowing a to be adjustable while b and c are fixed at b_o and c_o; (2°) in Eq. 9 the further reduction that is accomplished by allowing b to be adjustable while c is held at c_o; and (3°) in Eq. 10 the final reduction that comes from allowing c to be adjustable, the net result being the minimized sum of weighted squares, ϕ^2.

After a solution has been carried out upon the parameters a, b, c, the question often arises, what would have been the result for ϕ^2 if the parameter c had not been adjusted, but had been fixed at (say) γ_o? Now if this γ_o is not too far from the final value of c, one need only add $[cc.2](c-\gamma_o)^2$ to ϕ^2 in order to see what would have been obtained for ϕ^2 had c been fixed at γ_o (see examples 1 and 2 of section 20). The value of $\sigma^2(\text{ext})$ would then be $\phi^2 + [cc.2](c-\gamma_o)^2$ divided by n-2, not n-3 (n = the number of points).

Under certain conditions, the restriction that γ_o and c be not far apart can be removed; the polynomial $y = a + bx + cx^2$ with x free of error is an example. It all depends on whether the parameter c enters the L factors of tables 1 and 2 in section 16; if it does not, then no matter how wide the disparity between c and γ_o, the term $[cc.2](c-\gamma_o)^2$ still represents the increment in ϕ^2 that would be brought about by adjusting a and b to the condition $c = \gamma_o$.

In like manner, and under similar restrictions, a term $(b-\beta_o)^2/c_{22}$ will represent the increment in ϕ^2 that would be brought about by adjusting a and c to the condition $b = \beta_o$ (see exercise 2d).

Similarly, the two terms $[bb.1](b-\beta_o)^2$ and $[cc.2](c-\gamma_o)^2$ added to the ϕ^2 found in IV will give what would have been obtained for ϕ^2 if only a had been adjusted, b and c fixed at β_o and γ_o. The same result comes by subtracting $[aa](a-a_o)^2$ from [oo]. In this circumstance, $\sigma^2(\text{ext})$ would be computed with n-1 degrees of freedom.

It is important, as a practical matter, to note that the coefficients [bb.1] and [cc.2] needed for these increments are already at hand, numerically, in Eqs. II and III in the finished solution.

Remark 2 In both parts (a) and (c), ϕ^2 is shown as three terms subtracted from [oo]. Evidently

$$[ao]v_a + [bo]v_b + [co]v_c = [aa]v_a''^2 + [bb.1]v_b'^2 + [cc.2]v_c^2$$

The three terms of the right-hand member have individual significances, already discussed. The three terms on the left, however, so far as I see, have only collective significance; the three subtracted from [oo] give ϕ^2, but no one term represents a sum of weighted squares removed by a alone, b alone, or c alone. It might be helpful to turn to example 3 of section 20 at this point.

Exercise 4. Prove that the solution for v_a, v_b, and v_c found from the "1" column will also be given by the equations

$$\begin{aligned} v_a &= [ao]c_{11} + [bo]c_{12} + [co]c_{13} \\ v_b &= [ao]c_{21} + [bo]c_{22} + [co]c_{23} \\ v_c &= [ao]c_{31} + [bo]c_{32} + [co]c_{33} \end{aligned} \qquad (97)$$

This method of finding the unknowns v_a, v_b, and v_c is called the reciprocal solution because the reciprocal matrix is used as a multiplier along with the constant ("1") column [ao], [bo], [co]. The reciprocal solution is particularly useful when one has the same coefficients, hence the same reciprocal matrix, repeated over and over from one problem to another, but with a new constant column for each problem, and hence with a new set of values for v_a, v_b, and v_c each time. See, however, the reference to difficulties encountered in near indeterminacy, mentioned earlier in this section, also in example 1 of section 20. Theoretically, the direct and reciprocal solutions

should agree, and they will if the computer carries enough decimals.

Exercise 5. In matrix notation, the results of the last exercise are expressed by saying that we have

$$Av = H$$

H being the matrix of the "1" column, and that we desire to find the matrix v. The solution is evidently

$$v = A^{-1}H$$

To get A^{-1} we set

$$Ac = 1$$

and solve for c, getting

$$c = A^{-1}$$

Having now the matrix A^{-1}, we use it as a multiplier with H to find the matrix v from the relation above, getting v = cH. This is the matrix expression for the results stated in the preceding exercise. For illustrations see section 12d and example 1 of section 20.

Exercise 6. Prove that the values of the determinants A and A^{-1} are reciprocals.

Exercise 7. If $x = r \cos \theta$, $y = r \sin \theta$, the two Jacobians as matrices

$$\begin{vmatrix} \dfrac{dx}{dr} & \dfrac{dy}{dr} \\[2ex] \dfrac{dx}{d\theta} & \dfrac{dy}{d\theta} \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} \dfrac{\partial r}{\partial x} & \dfrac{\partial \theta}{\partial x} \\[2ex] \dfrac{\partial r}{\partial y} & \dfrac{\partial \theta}{\partial y} \end{vmatrix}$$

are reciprocals of one another; i.e., their product gives the unit matrix

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$$

18. The weights of the parameters; their standard errors. The standard error of a function of the parameters. The standard error of a curve. The confidence belts associated with a curve. It is a fact* that the reciprocals of the weights of the parameters are found on the diagonal of A^{-1} (see exercise 2 of the preceding section), i.e.

$$w_a = 1/c_{11}, \quad w_b = 1/c_{22}, \quad w_c = 1/c_{33} \tag{98}$$

Then since weights are reciprocals of variance coefficients (p. 12),

$$\sigma_a^2 = c_{11}\sigma^2, \quad \sigma_b^2 = c_{22}\sigma^2, \quad \sigma_c^2 = c_{33}\sigma^2 \tag{99}$$

$$\begin{aligned} (\text{S. E. of } a)^2 &= c_{11}\sigma^2 \\ (\text{"} \quad \text{"} \quad b)^2 &= c_{22}\sigma^2 \\ (\text{"} \quad \text{"} \quad c)^2 &= c_{33}\sigma^2 \end{aligned} \tag{100}$$

Let f be a function of the parameters. Then

$$\begin{aligned} \sigma_f^2 = (f_a\sigma_a)^2 &+ 2(f_a f_b r_{ab}\sigma_a\sigma_b + f_a f_b r_{ac}\sigma_a\sigma_c) \\ &+ (f_b\sigma_b)^2 + 2 f_b f_c r_{bc}\sigma_b\sigma_c \\ &+ (f_b\sigma_b)^2 \qquad \text{(by Eq. 47, p. 27)} \\ = \sigma^2 \{c_{11}f_a^2 &+ 2 c_{12}f_a f_b + 2 c_{13} f_a f_c \\ &+ c_{22}f_b^2 + 2 c_{23} f_b f_c \\ &+ 2 c_{33} f_c^2\} \end{aligned} \tag{101}$$

As in section 6c we write for the unbiased estimate of σ^2 by external consistency,

$$\sigma^2(\text{ext}) = \phi^2/k \quad \text{where } k = n - p$$

---

* The theory of all this goes back to Gauss, Theoria Combinationis (cited in section 6c), Art. 21. An excellent reference is to Whittaker and Robinson's Calculus of Observations Arts. 121-123.

n being the number of points and p the number of adjustable parameters. When σ is not known from any better source, this estimate may have to suffice, and $\sigma^2(\text{ext})$ would replace σ^2 in Eqs. 100 and 101, giving respectively

$$\begin{aligned}
(\text{Est'd S. E. of } a)^2 &= c_{11}\sigma^2(\text{ext}) \\
(\text{" " " } b)^2 &= c_{22}\sigma^2(\text{ext}) \qquad (102)\\
(\text{" " " } c)^2 &= c_{33}\sigma^2(\text{ext})
\end{aligned}$$

and

$$\begin{aligned}
(\text{Est'd S. E. of } f)^2 = \sigma^2(\text{ext}) \{c_{11}f_a &+ 2c_{12}f_af_b + 2c_{13}f_af_c \\
&+ 2c_{22}f_b^2 + 2c_{23}f_bf_c \\
&+ 2c_{33}f_c^2\} \qquad (103)
\end{aligned}$$

A convenient reference showing the application of this formula to curve fitting is a paper by Henry Schultz, J. Amer. Stat. Assoc. 25, 139-185, 1930. Shultz shows curves and confidence bands of width twice the S. E. of the curve for several kinds of curves. It should be mentioned, as Schultz does, that all these things were well known to Gauss and others in his time, but that they did not take the trouble to write out the formulas explicitly and draw the graphs for all the things that interest us today. It is a fact that of all the methods of curve fitting that have been devised, least squares is the only one for which the calculus of probabilities has been successfully applied toward the calculation of chances associated with the parameters and functions of them. The usefulness of Eqs. 102 or 103 or any like them depends only on whether the assumption of normally distributed observations is well enough obeyed, not on the intermediate steps of calculation, which are mathematical, and save for blunders, above reproach.

The student who has progressed thus far in the theory of least squares is strongly urged to study Ch. 5 of R. A. Fisher's Statistical Methods for Research Workers, wherein examples of the manipulation of the reciprocal matrix and the testing of parameters will be found, masterfully exhibited.

Only a brief discussion will be given here. When we write

$$y = f(x;\ a,b,c) \tag{104}$$

and ask for the S. E. of y, we are merely asking for the S. E. of a function of a, b, and c, but not of x; consequently, we can apply Eqs. 101 or 103 at once. This is what Schultz has done in his paper. x enters merely as a constant.

Confidence bands are of two kinds, (i) σ known, the normal type; (ii) σ estimated, the Student type. To construct a 50 percent confidence band of the Student type we look up t_{50} in Fisher's table of t for n-p degrees of freedom,* then compute $|Y-y|$ for several values of x, using the equation

$$t_{50} = \frac{|Y-y|}{\text{Est'd S. E. of } y} \tag{105}$$

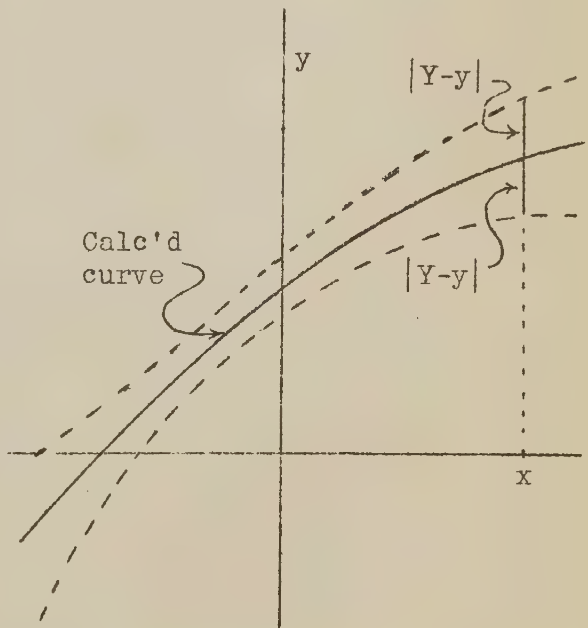

Fig. 11. A curve and confidence band

---

* V. A. Nekrassoff's very handy nomograph may be used in lieu of Fisher's table of t; it was published in Metron 8, No. 3, 1930, and is reproduced in W. A. Shewhart's Economic Control of Quality (Van Nostrand, 1931) p. 490; see also Deming and Birge Statistical Theory of Errors, p. 136.

The distance $|Y-y|$ laid off above and below the curve defines points on the confidence band. A 95 percent confidence band would be computed in the same way, but with t_{50} replaced by t_{95}; see Fig. 13 on page 168 for an example. (The capital Y used here is not to be confused with the same letter used in Fig. 9 and elsewhere for an observed coordinate).

The first type of confidence interval, the normal type, is constructed the same way, except that σ being known, the S. E. of y is known and not estimated, and the normal integral is used in place of the Student or t integral.

It must be remembered that a new set of points will give a new curve (i. e., a new set of parameters), hence if the confidence band is the normal type, the whole thing, curve and band, will be shifted to a new position by a new set of data. Moreover, in the Student type, the estimate of σ will fluctuate from one set of data to another; then not only will the curve and band be shifted to a new position by a new set of data, but the width of the band itself will also be different. The researches of Walter A. Shewhart have an important bearing on such matters; his Washington lectures (March 1938), to be published by the Graduate School, should be consulted. The reader is also referred to Neyman's Washington lectures (cited on page 18), pages 143-160, and to Fig. 11 on page 141 of Deming and Birge's Statistical Theory of Errors (cited on page 20).

Confidence intervals for any other function of a, b, and c are made up in like manner.

19. Exercises and notes on the formation of normal equations for various functions. The following examples are given as exercises for two purposes; first, to provide practice in setting up normal equations and gaining familiarity with the procedure in some common formulas; and second, in order to have the results for ready reference. Once these exercises are mastered, other forms of functions arising in practice should cause no difficulties.

A special note should be made concerning the fitting of polynomials, $y = a + bx$, $a + bx + cx^2$, and the like, x free of error. Some of the following exercises deal with this type of formula. Space will not be taken up here for a study of certain short-cuts that apply in the fitting of polynomials to equally spaced points (i.e. equally spaced along x), x free of error. To dismiss the subject with a reference, I shall mention what seems to be the most promising paper on the subject: Raymond T. Birge and John D. Shea "A rapid method for calculating the least squares solution of a polynomial of any degree", University of California Publications in Mathematics, vol. 2, No. 5, 1927, unfortunately now out of print. There is a strong possibility, however, that Professor Birge may publish a short account of the method with examples in the J. American Statistical Association for 1938 or 1939.

In this paper, Birge and Shea give explicit directions, tables, and examples for the fitting of polynomials. A feature of the method is its rapidity and ease, owing to the small number of figures required in the solution. Up to a certain point it seems to be equivalent to a method advocated by Harold T. Davis, but beyond that point the remaining steps are actually much simpler than Davis', and require only about half as many decimals. Any one having a great deal of work to do in the fitting of polynomials should become familiar with Birge and Shea's method. It is, of course, equivalent to--i.e. gives the same results as--any longer method that is truly a least squares solution.

Of late there have been many papers on the fitting of polynomials from the standpoint of orthogonal functions. The reader may wish to consult articles by A. C. Aitken, Proc. Royal Soc. Edinburgh 53; 54-78, 1932-33; also Max Sasuly's book Trend Analysis (The Brookings Institution, Washington, 1934). Another important contribution, using summation methods, is a paper by Frederick F. Stephan, J. Amer. Stat. Assoc. 27, 413-423, 1932.

Many people will prefer the form of solution given by R. A. Fisher in sections 28, 28.1 and 29.2 of the 6th edition of his Statistical Methods for Research Workers (1937). Miss Day tells me also that in the Forest Service, the method of Lorenz has been found very useful and easily learned; the reference is to Paul Lorenz Der Trend (Published for the Institut für Konjunkturforschung by Reimar Hobbing, Berlin S.W. 61, 1931). The copies are marked "Bezugpreis des Sonderhefts 14.40 RM."

THE LINE*

Exercise 1. (a) Given the line

$$y = a + bx$$

to be fitted to n points; x free of error, all y coordinates of equal weight (unity). Here we take

$$F = y - (a + bx)$$

The derivatives are $F_a = -1$, $F_b = -x$

$F_x = -b$ (not needed here), $F_y = 1$

$L = 1$ (Why? See Eq. 85, p. 85)

With the approximate values a_o and b_o we compute

$$F_o = y_{obs} - (a_o + b_ox)$$

at every point. Since L = 1 at every point, tables 1 and 2 of section 16 coalesce, and the normal equations are seen to be

No.	v_a	v_b	$\doteq$ 1	C_1	C_2	Sum
1	n	$[x]$	$-[F_o]$	1	0	...
2		$[x^2]$	$-[xF_o]$	0	1	...
3			$[F_oF_o]$	0	0	...

(Set 1, Exercise 1)

$[x]$ means Σx, $[F_oF_o]$ means the sum of the squares of F_o, etc.

---

* The line y = bx, forced to pass through the origin (i.e. with a = 0), was discussed to some extent in section 7.

(b) The solution for v_a and v_b found by the routine of section 17 or any other method of solution, is

$$v_a = [F_o]/n - \bar{x}\, v_b$$

$$v_b = -\ \{[xF_o] + \bar{x}[F_o]\}/n\mu_2$$

where $$n\mu_2 = [x^2] - n\bar{x}^2$$

$n\mu_2$ is the second moment of the x coordinates about an axis parallel to Oy and passing through the centroid $\bar{x}$, $\bar{y}$.

(c) The adjusted values of a and b turn out to be

$$a = a_o - v_a = \bar{y} - b\bar{x}$$

$$b = b_o - v_b = \Sigma(x-\bar{x})(y-\bar{y})/n\mu_2 = ([xy] - n\bar{x}\bar{y})/n\mu_2$$

The fitted line therefore passes through the centroid $\bar{x}$, $\bar{y}$. <u>But note</u> that when there is error in both x and y coordinates at some or all of the observed points, the weights being such that w_x/w_y is not constant throughout, the line does <u>not</u> pass through the centroid, see Remark 1 in Ex. 4.

(d) The solution just found for a and b is the same as would have been found from the normal equations

No.	a	b	=	1	C_1	C_2	Sum
I	n	[x]		[y]	1	0	-
2		$[x^2]$		[xy]	0	1	-
3				$[y^2]$	0	0	-

(Set 2, Exercise 1)

That is to say, it is permissible here to take a_o and b_o both as zero, whereupon F_o is simply y_{obs}. The normal equations Set 2 then give a and b directly.

<u>Note</u> that F_o is required at every point for Set 1, but not for Set 2. However, in Set 1 it is only the residuals v_a and v_b that are to be solved for, the main part of the adjustment having already been taken up in a_o and b_o; the additional accuracy required in Set 2 necessitates more figures in the formation and solution of the equations, the unknowns being the full values of a and b; and this will

usually more than offset the time required for computing F_0. It therefore is usually advisable to find good approximations and use Set 1. The better the approximations, the fewer figures required. Birge and Shea make use of this principle in their method of fitting polynomials (mentioned in the opening paragraphs of this section).

(e) When the solution of either Set 1 or 2 is carried out according to the scheme of calculation exhibited in section 17, the left-most entry in Eq. III will be the minimized $\mathcal{S}^2$ or $\Sigma(y_{obs} - y_{calc})^2$. The sum of squares removed from $[F_0F_0]$ in the case of Set 1, and from $[y^2]$ in the case of Set 2, by the successive adjustments of a and b, appear in the left-most entries of Eqs. 5 and 6. Show that

$$\mathcal{S}^2 = [yy] - n\bar{y}^2 - n\mu_2 b^2$$

the last term being the sum of squares removed by allowing the line to have slope b instead of slope 0--in other words, the <u>sum of squares removed by regression</u>. The two terms $n\bar{y}^2$ and $n\mu_2 b^2$ appear in Eqs. 5 and 6 of the solution of Set 2.

(f) If V denotes $y_{obs} - y_{calc}$ at any point, the solution for a and b renders $\Sigma V = 0$. (But note that neither ΣV nor ΣwV is, in general, zero in least squares solutions; it only happens to be so here. In fact, in section 7a we saw a simple example wherein neither ΣV nor ΣwV was zero. See remark 3 in exercise 4; also exercise 5).

<u>Exercise 2</u>. (a) The reciprocal matrix for the normal equations in the preceding exercise appears in the C_1 and C_2 columns of Eqs. 7 and 8 (these numbers refer to the solution of the equations solved according to the form in section 17). It turns out to be

$$A^{-1} = \begin{vmatrix} \dfrac{1}{n} + \dfrac{\bar{x}^2}{n\mu_2} & -\dfrac{\bar{x}}{n\mu_2} \\ -\dfrac{\bar{x}}{n\mu_2} & \dfrac{1}{n\mu_2} \end{vmatrix}$$

(b) Hence the weight of a is $n/(1+\bar{x}^2/\mu_2)$, and the weight of b is $n\mu_2$ (see section 18). Our confidence in b thus increases as the "spread" of the points increases. Is this reasonable? Why does the weight of a depend on $\bar{x}$?

(c) The
$$(\text{S.E. of } a)^2 = \sigma^2(1/n + \bar{x}^2/n\mu_2)$$
$$(\text{S.E. of } b)^2 = \sigma^2/n\mu_2$$

Note that the weights and S.Es. of a and b do not involve the y coordinates of the points. Compare with part (d) of the next exercise. Note also that

$$n^2\mu_2 = \begin{vmatrix} n & [x] \\ [x] & [x^2] \end{vmatrix} = A;$$

hence near-indeterminacy (small A) is closely associated with high S.Es. of a and b, and a rapid "fanning out" of the S.E. of y_{calc} each side of $\bar{x}$, $\bar{y}$ (see the next part; also exercise 2a of section 17).

(d) From Eq. 48, p. 27 and the reciprocal matrix of part (a), prove that the

$$(\text{S.E. of } y_{calc})^2 = \frac{\sigma^2}{n}\{1 + (x-\bar{x})^2/\mu_2\}$$

Thus the S.E. of y_{calc} is least at the centre of gravity ($\bar{x}$, $\bar{y}$) of the points, and fans out each side of it. See section 18 and reference to Henry Schultz; also Figs. 11 and 12.

(e) The S.E. of the calculated line of Ex. 1 at the centre of gravity is $\sigma/\sqrt{n}$, as it would be for n observations made on a single unknown. (Do this in two ways: 1° put $x = \bar{x}$ in part d; 2° put $\bar{x} = 0$ in part c for the S.E. of a).

Exercise 3. (a) Carry out the solution of the normal equations of Ex. 1a in symbols, following the outline given in section 17, and show that the minimized value of ϕ^2 or of $\Sigma(y_{obs} - y_{calc})^2$ comes in the "1" column of Eq. III (which will be the left-most entry in III). The same is true if the approximations a_o and b_o are used, as is advised in Ex. 1d.

(b) Show that this can be written

$$\Sigma \text{res}^2 = n(1-r^2)s_y^2$$

where $r = \Sigma(x-\bar{x})(y-\bar{y})/ns_xs_y$ = the correlation coefficient

and $ns_y^2 = [yy] - n\bar{y}^2$

$ns_x^2 = [xx] - n\bar{x}^2$

(s_x^2 is here used in place of μ_2 for consistency with s_y).

(c) The estimate of σ made from the fit of the line is (see section 6c)

$$\sigma^2(\mathrm{ext}) = n(1-r^2)s_y^2/(n-2)$$

(d) The $(\text{Est'd S.E. of } a)^2 = \dfrac{1-r^2}{n-2}\, s_y^2\,(1+\bar{x}^2/s_x^2)$

$$(\text{Est'd S.E. of } b)^2 = \frac{1-r^2}{n-2}\,(s_y^2/s_x^2)$$

Note that the Est'd S.Es. of a and b involve the y coordinates: compare with part (c) of the preceding exercise.

(e) The $(\text{Est'd S.E. of } y_{calc})^2 = \dfrac{1-r^2}{n-2}\, s_y^2\,\{1+(x-\bar{x})^2/s_x^2\}$

Exercise 4 (a) If both x and y coordinates are subject to error with varying precisions at some or all of the n points, one must go through the calculation of tables 1 and 2 of section 16. For the line

$$y = a + bx$$

one takes

$$F \equiv y - a - bx$$

Some good approximate values a_o and b_o having been found, one can then calculate the numerical values of

$$F_o = y_{obs} - (a_o + b_o x_{obs})$$

at the n points.

The derivatives of F are

$$F_x = b, \quad F_y = 1, \quad F_a = -1, \quad F_b = -x$$

whereupon $$L = b^2/w_x + 1/w_y$$

L varies from point to point with w_x or w_y.

[Sec. 19,
Ex. 4, the line]

The headings for table 1 of section 16 would be

h, or Point No.	w_x	w_y	L	$\surd L$	$F_b = -x$	F_o

(It is not necessary to tabulate F_x, F_y, F_a, and F_b since they remain constant from point to point).

The headings for table 2 of section 16 are then

h, or Point No.	$F_a/\surd L = -1/\surd L$	$F_b/\surd L = -x/\surd L$	$F_o/\surd L$	Sum

It has already been remarked that there is some theoretical advantage in writing W in place of 1/L, though it is a fact that with machines having automatic division and two dials for quotients--one for the individual quotients needed for table 2, and another for cumulating the quotients across the rows for the "Sum" column of table 2--there is a practical advantage in tabulating $\surd L$ rather than $\surd W$ in table 1, and using divisions by $\surd L$, rather than multiplications by $\surd W$, to form table 2.

Writing now

$$1/W = b^2/w_x + 1/w_y$$

we see that $W = w_y$ if x is free of error or if b = 0 (see the next exercise), and $W \doteq w_x/b^2$ if y is free of error, but that both terms are required if x and y are both subject to varying errors (see Ex. 8b).

In terms of W the headings of table 2 would be

h, or Point No.	$-\surd W$	$-x\surd W$	$\surd W F_o$	Sum

(b) The normal equations are formed in the usual way by summing squares and cross-products from table 2; they can be symbolized as

No.	v_a	v_b	= 1	C_1	C_2	Sum
I	[W]	[Wx]	$-[WF_o]$	1	0	-
2		$[Wx^2]$	$-[WxF_o]$	0	1	-
3			$[WF_oF_o]$	0	0	-

(Set 1, Exercise 4)

The solution of the normal equations gives v_a and v_b from the "1" column, and the reciprocal matrix A^{-1} as usual from the C_1 and C_2 columns. The adjusted values of a and b will be

$$a = a_0 - v_a$$

$$b = b_0 - v_b$$

The solution gives the minimized value of $\sum(w_x V_x^2 + w_y V_y^2)$ in Eq. III, column "1". That portion of the sum of the weighted squares subtracted from $[WF_0F_0]$ by shifting the first parameter from a_0 to a appears just above in Eq. 5, and the portion further removed by shifting the second parameter from b_0 to b appears in Eq. 6 (see exercise 3 of section 17; also exercise 1 of this section).

Note that a_0 and b_0 may be taken as 0 (with the necessary increase in the number of decimals required in the normal equations), so far as F_0 is concerned, in which event F_0 becomes simply

$$F_0 = y_{obs}$$

The normal equations are then written

No.	a	b	=	1	C_1	C_2	Sum	
I	[W]	[Wx]		+[Wy]	1	0	-	
2		[Wx²]		+[Wxy]	0	1	-	(Set 2, Exercise 4)
3				[Wy²]	0	0	-	

and the solution gives a and b directly. Why are more decimals required in these equations than in the preceding ones giving the (supposedly small) residuals v_a and v_b?

But note carefully that an approximate value of b must be used in the calculation of W at each point where x is subject to error. b_0 may be called 0 in the calculation of F_0 (as noted above) but not in the calculation of W. If this admonition is disregarded, the effect of the weighting of x is lost. In fact, if it turns out that the approximate value of b used for calculating W was too far removed from the final b, it may be desirable to make a second adjustment to secure improved weightings W, which can be obtained by using the value of b from the first adjustment; but this is seldom found necessary in practice.*

---

* As Gauss put it, in a somewhat different problem: "Quodsi dein calculo absoluto contra exspectationem valores incognitarum p', q',

Of course, it may happen in some particular problem that b actually is very small, and that 0 is therefore a good approximation for b. In such circumstances the line is practically horizontal, and the weighting of x does not matter much, and the computer may as well simplify matters and set $W = w_y$, ignoring the weighting of x--not because x is free of error (i.e. not because w_x is infinite), but because b is zero or nearly so. The student should ponder over the situation where b is actually known to be 0; do the values of x count at all in the solution? Does this not take us back to "the simplest problem in least squares" of section 3, the x and y axes being interchanged? The solution would be simply

$$a = \Sigma wy/\Sigma w$$

and the weight of a would be Σw, its S.E. $\sigma/\Sigma w$ (w being written for w_y for convenience).

<u>Remark 1</u> When x and y are both subject to error at some or all of the observed points, the line does not pass through the centre of gravity

$$\bar{x} = [xw_x]/[w_x], \qquad \bar{y} = [yw_y]/[w_y]$$

But the line will in any case pass through a <u>quasi centre</u> defined as

$$x' = [xW]/[W] , \qquad y' = [yW]/[W]$$

<u>Remark 2</u> With 1/W written in place of L, Eq. 85, p. 85 gives

$$\frac{1}{W} = \frac{F_xF_x}{w_x} + \frac{F_yF_y}{w_y}$$

As has already been seen, the first term drops out if x is free of error, and the second term drops out if y is free of error. To make the change from a solution in which x is free of error to one wherein both coordinates are subject to error, we merely add the other term in 1/W and recalculate W at every point, the procedure being otherwise the same. There is a close analogy with celestial mechanics; when one wishes to compute the orbit of a body of mass m about another of mass M, he may at first make the simplifying assumption that M is infinite

---

r', s' etc. tanti emergerent, ut parum tutum videatur, quadrata productaque neglexisse, eiusdem operationis repititio (acceptis loco ipsarum π, χ, ρ, σ etc. valoribus correctis ipsarum p, q, r, s, etc.) remedium promtum afferet." <u>Theoria Motus Corporum Coelestium</u> (Hamburg, 1809) Art. 180.

(i.e. immovable), and solve the equations, later replacing m by μ where

$$\frac{1}{\mu} = \frac{1}{m} + \frac{1}{M}$$

This replacement yields the absolute motion of the two bodies, neither being of infinite mass (i.e. neither one immovable).

Remark 3 When x and y are both subject to error at some or all of the points, we can not always assert that

$$\Sigma V_x = 0, \text{ or } \Sigma V_y = 0, \text{ or } \Sigma(w_x V_x + w_y V_y) = 0$$

though these may sometimes happen, as in Exs. 1 and 5, q. v. We have already seen a simple example in section 7a where these summations were not zero. There is, however, a property of least squares by which one can always assert that after the adjustment,

$$\Sigma(w_x U_x V_x + w_y U_y V_y) = 0$$

(For definitions of U_x, etc., see Figs. 8 and 9 on pp. 82 and 83. This property of least squares was published by me in the Phil. Mag. 19, 389-402, 1935).

Remark 4. It is interesting to note that in the routine solution of Set 2, the minimized ϕ^2 appears in the left-most entry of Eq. III, but that, in contrast with Set 1, unless the final value of b is actually 0 or very near, the entry in Eq. 6 directly above ϕ^2 will not show the increment in ϕ^2 that would result from fixing b at either the value 0 or b_0. The reason is that a good value of b must be used in W at each point where x is subject to error: if we want to know what the solution would have been with b = 0, we must actually make a solution with b set equal to 0 in the computation of W, in which circumstance W reduces to w_y, as already noted.

Exercise 5 . Given

$$y = a + bx$$

to be fitted to n points, x free of error, the y coordinate having a weight w_y, varying from point to point. This is similar to Ex. 1 except that now the y coordinates have unequal precisions. Here we take

$$F = y - a - bx$$

as in the preceding exercise, the derivatives being also the same. But since x is free of error, w_x is infinite and it follows that W (defined as 1/L) is none other than w_y. All we have to do is replace W in the preceding exercise by w_y, and the results will apply here. The headings of table 1 of section 16 will be

h, or Point No.	w_y	$\sqrt{w_y}$	$F_b = -x$	F_o

For table 2

h, or Point No.	$\sqrt{w_y}$	$F_b/\sqrt{L} = -x\sqrt{w_y}$	$\sqrt{w_y}F_o$	Sum

The normal equations are written in the same symbols as those of the preceding exercise, but with w_y in place of W. Eq. III in the solution of the normal equations gives the minimized value of $\Sigma w_y(y_{obs} - y_{calc})^2$ In Eqs. 5 and 6 are found the portions of the weighted squares removed by a and b, as in exercise 1e.

Note that as in the preceding exercise, it is permissible to take a_o and b_o as zeros, if the number of decimals in the normal equations is increased accordingly. In this event,

$$F_o = y_{obs}$$

and the normal equations may be written

No.	a	b	= 1	C_1	C_2	Sum
I	[w]	[wx]	[wy]	1	0	-
2		[wxx]	[wxy]	0	1	-
3			[wyy]	0	0	-

giving a and b directly. Eq. III will give the minimized value of $\Sigma w_y(y_{obs} - y_{calc})^2$ in the "1" column and Eqs. 5 and 6 will show the sum of squares removed successively by a and b, as in exercise 1e; since w_x is infinite, the question of an approximate value of b for use in the calculation of W does not come up.

Remark For the conditions stated (x free of error) the sum of the weighted y-residuals is zero, i.e.

$$\Sigma wV = 0$$

(See remark 3 of exercise 4)

Exercise 6. (a) Given the line

$$y = a + bx$$

to be fitted to n points, both x and y coordinates subject to error but in such a way that w_x/w_y is constant and not infinite nor zero, the line passes through the centre of gravity $\bar{x} = [w_x x]/[w_x]$, $\bar{y} = [w_y y]/[w_y]$, with slope

$$b = \frac{c[wv^2] - [wu^2] + \sqrt{\{c[wv^2] - [wu^2]\}^2 + 4c[wuv]^2}}{2c[wuv]}$$

(This is equivalent to a result obtained by Kummell in 1876, Karl Pearson in 1901, and Gini in 1921; references in section 1). Here u and v are the x and y coordinates of a point, measured from the center of gravity $\bar{x}$, $\bar{y}$; i.e. $u_i = x_i - \bar{x}$, and $v_i = y_i - \bar{y}$. c is written for w_y/w_x, and w in place of w_x, for convenience.

(b) If the plus sign be changed to minus in front of the radical, the result is the slope of the worst fitting line, that which maximizes the value of $\Sigma(w_x V_x^2 + w_y V_y^2)$.

(c) Prove that under these conditions of weighting, the best and worst fitting lines are perpendicular to each other.

Exercise 7. (a) Given

$$y = a + bx$$

to be fitted to n points when y is free of error and all x coordinates are of equal weight (unity), we may write

$$x = p + qy$$

and find the following normal equations for p and q.

No.	p	q	=	1	C_1	C_2	Sum
I	n	[y]		[x]	1	0	-
2		[yy]		[yx]	0	1	-
3				[xx]	0	0	-

which are like Set 2 of exercise 1 with x and y interchanged.

Eq. III in the solution of the normal equations gives Σ res^2 where now the deviations are measured parallel to the x axis.

(b) The reciprocal matrix A^{-1} found in the C_1 and C_2 columns of Eqs. 7 and 8 of the solution will be

$$A^{-1} = \begin{vmatrix} 1/n + \bar{y}^2/ns_y^2 & -\bar{y}/ns_y^2 \\ -\bar{y}/ns_y^2 & 1/ns_y^2 \end{vmatrix}$$

where s_y^2 has the same significance as in exercise 3.

(c) (The S.E. of $x_{calc})^2 = (\sigma^2/n)\{1 + (y-\bar{y})^2/s_y^2\}$

(d) The normal equations of exercise 7a give $\Sigma V_x = 0$.

(See remarks in exercises 1f, 4, and 5).

Exercise 8. (a) Prove that with y free of error, and all x coordinates of equal precision, the normal equations for a and b (or for v_a and v_b) in exercise 4 will give the same line as the normal equations in exercise 7a (i.e. will give $p = -a/b$ and $q = 1/b$), except for the effect of the neglect of the squares of the x-residuals. The solution of exercise 7a is the more accurate in not throwing away any higher powers of residuals. This may occasionally be important. See also exercises 18 and 22.

(b) Show that if x has the same weight (i.e. the same precision) over all n points, and y likewise, x and y both subject to error, the line that one gets by the exact solution given in exercise 6a lies between the two false lines that one gets by (1°) throwing the adjustment all on to y, using the equations of exercise 1; and (2°) throwing the adjustment all on to x, using the equations of exercise 7a; but that these two false lines differ only in the effect of the squares of the x-residuals and of the y-residuals, respectively. (Hint: both terms of 1/W in exercise 4 are constant over all points when w_x and w_y are constant; hence, so far as the values of a and b or p and q are concerned, W can be put equal to unity at every point in all three solutions--in the correct solution, and in the two false solutions. The normal

equations of exercise 4 will then give identical results for all three. But the normal equations of exercise 4 can be in error at most by the neglect of higher powers of the residuals, hence the false solutions 1° and 2° can differ from the true solution only through the neglect of such terms.. This means that when x has the same weight over all points, and y likewise, the false solutions will hardly be distinguishable from the true solution if the residuals are all fairly small.*

Note that if W is constant from one point to another, it is advisable for convenience of computation to choose the system of weighting so that $W = 1$ at all points. This is only saying that the arbitrary factor ϕ^2 in Eq. 13 of article 5 is to be chosen so that $W = 1$. Then σ^2 in the left-most entry of Eq. III in the solution of the normal equations comes out in the same system, and $\sigma^2(\text{ext}) = \phi^2/(n-2)$ is the external estimate of σ^2 in the same units as were arbitrarily chosen for it.

(c) All three lines of part (b) pass through the centre of gravity $\bar{x}$, $\bar{y}$ (called also the centroid).

Remark. Statements similar to those of part (b) will hold for any curve when the combination of the form of the function and the weighting of the coordinates causes both terms of $1/W$ to be constant over all n points. Example 3 of the next section is an illustration in three dimensions (three terms in $1/W$).

Exercise 9. For the line $y = a + bx$ fitted to n points, the following expressions hold (all due to Karl Pearson, Phil. Mag. (London) 2, 559-572, 1901).

(a) $\Sigma\, \text{res}^2 = n(1-r^2)s_y^2$ — When x is free of error, the y coordinates all of equal weight (unity); the deviations measured in the vertical. (This result was given in a previous exercise).

(b) $\Sigma\, \text{res}^2 = n(1-r^2)s_x^2$ — When y is free of error, the x coordinates all of equal weight (unity); the deviations measured in the horizontal.

---

* This fact was noted by me without proof in the Proc. Physical Soc. (London) 47, p. 107, 1935.

(c) $\Sigma \text{res}^2 = \frac{1}{2}n\{s_x^2 + s_y^2 - \sqrt{(s_x^2 - s_y^2)^2 + 4r^2 s_x^2 s_y^2}\}$

When the x and y coordinates are of equal weight (unity), the deviations measured perpendicular to the fitted line.

In these formulas, s_x^2, s_y^2, and r have the meaning ascribed to them in exercise 3.

THE PARABOLA

Exercise 10. Given

$$y = a + bx + cx^2$$

to be fitted to n points, x and y having weights w_x and w_y at any point. Here we take

$$F = y -(a + bx + cx^2)$$

$$F_o = y_{obs} -(a_o + b_o x_{obs} + c_o x_{obs}^2) \text{ as usual.}$$

The derivatives of F are

$$F_x = -(b + 2cx), \qquad F_y = 1$$

$$F_a = -1, \qquad F_b = -x, \qquad F_c = -x^2$$

$$L \text{ or } 1/W = (b + 2cx)^2/w_x + 1/w_y$$

The headings of table 1 in section 16 will be

h, or Point No.	$F_x = -(b + 2cx)$	w_x	w_y	L	$\sqrt{L}$	F_b	F_c	F_o

It is understood that in calculating all quantities under these headings, x and y are to be replaced by their observed values, and a, b, c, by a_o, b_o, c_o. It is not necessary to tabulate F_y and F_a because they remain constant from point to point.

The headings of table 2 will be as follows

h, or Point No.	$F_a/\sqrt{L}$ $= -\sqrt{W}$	$F_b/\sqrt{L}$ $= -\sqrt{W}x$	$F_c/\sqrt{L}$ $= -\sqrt{W}x^2$	$F_o/\sqrt{L}$ $= \sqrt{W}F_o$	Sum

The usual process of cumulating sums of squares and cross-products in table 2 yields the normal equations

No.	v_a	v_b	v_c	= 1	C_1	C_2	C_3	Sum	
I	[W]	[Wx]	$[Wx^2]$	$-[WF_o]$	1	0	0	-	
2		$[Wx^2]$	$[Wx^3]$	$-[WxF_o]$	0	1	0	-	(Set 1,
3			$[Wx^4]$	$-[Wx^2F_o]$	0	0	1	-	Ex.10)
4				$-[WF_oF_o]$	0	0	0	-	

The solution, carried out by the usual routine procedure, gives v_a, v_b, and v_c, whence the adjusted values of a, b, and c are

$$a = a_o - v_a$$

$$b = b_o - v_b$$

$$c = c_o - v_c$$

The minimized value of ϕ^2 or $\Sigma(w_xV_x^2 + w_yV_y^2)$ will appear in Eq. IV, column "1". This will be simply $\Sigma\, w_xV_x^2$ if y is free of error, and $\Sigma\, w_yV_y^2$ if x is free of error. Directly above, in Eq. 8, appears the sum of squares that is removed from $[WF_oF_o]$ by shifting the y-intercept from a_o to a; in Eq. 9 appears the further decrease brought about by allowing the second parameter to shift from b_o to b; and in Eq. 10, just above ϕ^2, appears the portion of the sum of squares that is finally removed by adjusting the parabolic term from c_ox^2 to cx^2 (see Ex. 3 of sec. 17).

The reciprocal matrix A^{-1} will appear in the C_1, C_2, C_3 columns of Eqs. 11, 12, and 13 (the "back solution"), containing the variance and product-variance coefficients for a, b, and c. (See Ex. 12 for the matrix A^{-1} in a special case.)

Note the similarity between Sets 1 of exercises 4 and 10.

Note that if x is free of error, a_o, b_o, and c_o may be taken as 0 so far as F_o is concerned, in which event F_o becomes simply

$$F_o = y_{obs}$$

The normal equations giving a, b, and c directly are then

No.	a	b	c	= 1	C_1	C_2	C_3	Sum
I	[W]	[Wx]	[Wx²]	[Wy]	1	0	0	-
2		[Wx²]	[Wx³]	[Wxy]	0	1	0	-
3			[Wx⁴]	[Wx²y]	0	0	1	-
4				[Wyy]	0	0	0	-

(Set 2, Ex. 10)

More decimals will of course be required than if good values of a_o, b_o, and c_o had been used in the calculation of F_o, and the previous normal equations set up for v_a, v_b, and v_c. Why? The reciprocal matrix is unchanged, and the minimized value of $\Sigma(w_x V_x^2 + w_y V_y^2)$ again comes in the left-most entry of Eq. IV; but, as in exercise 4, remark 4, the entries directly above it in Eqs. 10 and 9 do not show the increments in the sum of the weighted squares that would result from setting c = 0 and b = c = 0, respectively. The only exceptions come when b and c are very small, or x free of error.

Note the similarity between Sets 2 of exercises 4 and 10.

Note carefully that approximate values of b and c must be used in the calculation of W at each point where x is subject to error. See the notes at the end of exercise 4, which apply here with obvious modifications.

Exercise 11. Given*

$$y = a + bx + cx^2$$

to be fitted to n points, x free of error, all y coordinates of equal weight (unity). If a_o, b_o, and c_o all be taken as 0, the normal

---

* See the notes at the beginning of this section for special methods applying when x is equally spaced.

equations giving a, b, and c directly are

No.	a	b	c	=	1	C_1	C_2	C_3	Sum
I	n	[x]	$[x^2]$		[y]	1	0	0	-
2		$[x^2]$	$[x^3]$		[xy]	0	1	0	-
3			$[x^4]$		$[x^2y]$	0	0	1	-
4					[yy]	0	0	0	-

These arise immediately from the last set of normal equations of the preceding exercise by noting that under the conditions $w_x = \infty$ and $w_y = 1$ throughout, W = 1 throughout, Eq. IV in the solution of the normal equations will contain the minimized value of $\Sigma(y_{obs} - y_{calc})^2$ in the "1" column. The sum of squares successively removed by the constant, linear, and parabolic terms will appear in the "1" column of Eqs. 8, 9, and 10 (see exercise 3 of section 17; also exercises 1, 4, 5, and 10 of this section).

Note the similarity between these normal equations and those of exercise 1, Set 2.

Exercise 12. (a) In the preceding exercise, let the origin of x be taken at the mean value of x, and let $n\mu_2$ be written for $[x^2]$ and $n\mu_4$ for $[x^4]$. Then the normal equations are

No.	a	b	c	=	1	C_1	C_2	C_3
I	n	0	$n\mu_2$		[y]	1	0	0
2		$n\mu_2$	0		[xy]	0	1	0
3			$n\mu_4$		$[x^2y]$	0	0	1
4					[yy]	0	0	0

Find the reciprocal matrix to be

$$A^{-1} = \begin{vmatrix} \dfrac{\mu_4}{n\mu_4 - n\mu_2^2} & 0 & -\dfrac{\mu_2}{n\mu_4 - n\mu_2^2} \\ 0 & \dfrac{1}{n\mu_2} & 0 \\ -\dfrac{\mu_2}{n\mu_4 - n\mu_2^2} & & \dfrac{1}{n\mu_4 - n\mu_2^2} \end{vmatrix}$$

This enables one to write down the S. E. of the fitted curve, or any function of a, b, c, in terms of σ (see section 8 and exercise 2 of this section). In particular,

$$w_a = n(1 - \mu_2^2/\mu_4)$$

$$w_b = n\mu_2, \text{ as in Ex. 2}$$

$$w_c = n(\mu_4 - \mu_2^2)$$

The (S. E. of y_{calc})^2 at the centre of gravity is equal to $\sigma^2/n(1 - \mu_2^2/\mu_4)$, which exceeds the value $\sigma^2 n$ which was found in exercise 2 for the line.

(b) Show that the determinant A of the coefficients is equal to $n^3\mu_2(\mu_4 - \mu_2^2)$; hence that near-indeterminancy (small A) will result not only in instability but also in high S.Es. for a and b and rapid fanning out of the S. E. of y_{calc} (see exercise 2c; also exercise 2a of section 17).

THE EXPONENTIAL AND ITS LOGARITHMIC FORM

Exercise 13. Given the equation

$$y = ae^{bx}$$

to be fitted to n points, x free of error, all y coordinates of equal precision (unit weight). Here we take

$$F = y - ae^{bx}$$

Good approximate values of a and b can usually be found by plotting log y against x: assuming that they can be obtained, we write

$$F_o = Y - a_o e^{b_o x}$$ (Y denotes an observed y as in Fig. 9, p. 83)

The derivatives of F are

$$F_y = 1, \quad F_a = - y/a, \quad F_b = - xy$$

W = 1 at all points, hence tables 1 and 2 of section 16 coalesce. They will be made up as follows. For convenience in writing, the subscript o will be withheld from the a and b. X_1, Y_1, etc., are the observed x and y coordinates of the n points.

Tables 1 and 2

h	F_a	F_b	F_o	Sum
1	$-Y_1/a$	$-X_1Y_1$	$Y_1 - ae^{bx_1}$	-
2	$-Y_2/a$	$-X_2Y_2$	$Y_2 - ae^{bx_2}$	-
⋮	⋮	⋮	⋮	
etc.				

The normal equations are formed from this table by summing squares and cross products:

No.	v_a	v_b	=	1	C_1	C_2	Sum
I	$[Y^2/a^2]$	$[XY^2/a]$		$-[YF_o/a]$	1	0	-
2		$[X^2Y^2]$		$-[XYF_o]$	0	1	-
3				$[F_oF_o]$	0	0	-

The solution of the normal equations by the usual routine described in section 17 gives v_a and v_b, also the reciprocal matrix, and the minimized value of $\sum res^2$. The adjusted values of a and b are then

$$a = a_o - v_a$$

$$b = b_o - v_b$$

The left-most entry in Eq. III of the solution of the normal equations gives ϕ^2 or $\sum res^2$, the deviations all being measured in the vertical (i. e. parallel to Oy). Directly above ϕ^2, in Eq. 5, appears the sum of squares removed by the shift from a_o to a, and in Eq. 6 the further decrease brought about by adjusting the exponent from $b_o x$ to bx.

Exercise 14. If in the preceding exercise, the x coordinates are free of error but the y coordinates have unequal precision, designated by weight w (varying from point to point), W is no longer unity, but is equal to w. Table 2 of section 16 then runs as follows:

h	$F_a/\sqrt{L}$	$F_b/\sqrt{L}$	$\sqrt{w}F_o$	Sum
1	$-\sqrt{w_1}Y_1/a$	$-X_1Y_1\sqrt{w_1}$	$\sqrt{w_1}(Y_1-ae^{bx_1})$	-
2	$-\sqrt{w_2}Y_2/a$	$-X_2Y_2\sqrt{w_2}$	$\sqrt{w_2}(Y_2-ae^{bx_2})$	-
⋮	⋮	⋮	⋮	
etc.				

the approximate values a_o and b_o being inserted for a and b.

The normal equations, formed from table 2 in the usual manner, can be symbolized as follows:

No.	v_a	v_b	= 1	C_1	C_2	Sum
I	$[wY^2/a^2]$	$[wXY^2/a]$	$-[wYF_o/a]$	1	0	-
2		$[wX^2Y^2]$	$-[wXYF_o]$	0	1	-
3			$[wF_oF_o]$	0	0	-

In the solution of the normal equations by the routine of section 17, the minimized value of $\Sigma w(y_{obs} - y_{calc})^2$ comes in the left-most entry of Eq. III. Just above, in Eqs. 5 and 6, will appear the portions of the weighted sums of squares removed successively by adjusting a and then b (see exercise 3 of section 17; also exercises 1, 4, 5, 10, 11, 12, and 13 of this section).

The adjusted values of a and b are, as usual,

$$a = a_o - v_a$$

$$b = b_o - v_b$$

v_a and v_b being found by solving the normal equations. Naturally these normal equations become the same as those of exercise 13 if w = 1 throughout.

Exercise 15. (a) The formula to be fitted in the two previous exercises can be taken in the logarithmic form

$$\log y = \log a + bx \log e$$

Suppose now that y' be written for log y, a' for log a, x' for log x, B for b log e; then

$$y' = a' + Bx$$

We now take

$$f = y' - (a' + Bx)$$

$$f_o = Y' - (a_o' + B_o x) \qquad (Y' = \log Y_{obs})$$

wherein a_o' means log a_o, and B_o means b_o log e. The derivatives of f are as follows:*

$$f_x = -B, \qquad f_{y'} = 1, \qquad f_y = \frac{df}{dy'}\frac{dy'}{dy} = 0.434/y$$

$$f_{a'} = -1, \qquad f_B = -x$$

$$L \text{ or } 1/W = F_x F_x/w_x + F_y F_y/w_y = B^2/w_x + 0.434^2/y^2 w_y$$

---

* It is convenient to remember that log e = 1/ln 10 = 0.434 ... = 1/2.30 ... log means base 10, ln means base e (logarithme naturel).

Suppose that x is free of error, but that the weight of an observed y coordinate is w, which may vary from point to point. (If all y coordinates have equal weight, it is easy to put w equal to 1 in what follows). With $w_x = \infty$, the first term in L drops out, leaving

$$L \text{ or } 1/W = 0.434^2/y^2w$$

It will be noticed from the result of exercise 8c, page 30, that if w is the weight of y, then $y^2w/0.434^2$ is the weight of y' or log y. Suppose that on this account we set

$$w' = y^2w/0.434^2 = (2.30y)^2w \qquad \text{(See p. 30)}$$

then w' is the weight of y', and

$$W = w'$$

If w (the weight of y) is constant throughout, then w'_y is not, and vice versa. (Cf. the remark appended to exercise 18).

Table 2 of section 16 will have headings as follows:

h, or Point No.	$f_{a'}/\sqrt{L} = -\sqrt{w'}$	$f_B/\sqrt{L} = -x\sqrt{w'}$	$\sqrt{w'}f_o$	Sum

The normal equations will be

No.	$v_{a'}$	v_B	1	C_1	C_2	Sum
I	[w']	[xw']	$-[w'f_o]$	1	0	-
2		$[x^2w']$	$-[xw'f_o]$	0	1	-
3			$[w'f_of_o]$	0	0	-

We perceive that these are similar to the normal equations of exercise 5, but with w' in place of w, and f in place of F. More precisely, the comparison is this:

In Ex. 5, y = a + bx, x free of error, y of weight w.

Here, y' = a' + Bx, " " " ", y' " " w'.

This means that we may fit the equation $y = ae^{bx}$ by writing it in the logarithmic form

$$\log y = \log a + bx \log e$$

and treating it as a linear equation in log y and x, at the same time giving ln y a weight just y^2 times the weight of y, or log y a weight $(2.30\ y)^2$ times the weight of y.

Remark 1 It is customary among computers to fit the exponential equation $y = ae^{bx}$ by taking logarithms and treating it as linear in log y and x, but it is not so usual for them to change the weighting to correspond to the logarithms. The neglect of the factor $(y\ 1.434)^2$ not only produces somewhat incorrect results for a and b, but also invalidates the reciprocal matrix and all calculations made with it on the S. E. of a function of the parameters; moreover, under such circumstances, the left-most entry of Eq. III no longer contains ϕ^2. In fact, all calculations are vitiated, the spurious values for a and b so obtained being about the least objectionable feature. See also some remarks at the conclusion of exercise 18.

(b) The left-most entry of Eq. III in the solution of the normal equations contains $\Sigma w(y_{obs} - y_{calc})^2$. The left-most numbers appearing in Eqs. 5 and 6 are the weighted sums of squares removed successively by adjusting a and then b (see exercise 3 of section 17 and exercises 1, 4, 5, 10-14 of this section). These statements would not be true if one were to neglect the factor $(2.30\ y)^2$ for the weight of log y.

Note that in the normal equations of part (a) it is permissible to use $a_0' = 0$ and $B_0 = 0$, in which event

$$f_0 = Y' \text{ or } \log Y_{obs}$$

whereupon the normal equations will be

No.	a'	B	1	C_1	C_2	Sum
I	[w']	[w'x]	[w'Y]	1	0	-
2		[w'x²]	[w'xY]	0	1	-
3			[w'YY]	0	0	-

($w' = 2.30^2\ y^2\ w$ as on the preceding page)

giving a' and B directly. As in exercises 1 and 5, no question of an approximate value of B enters for the calculation of the w', since w_x is infinite (x free of error), but more decimals are required than when good approximate values of a and b are used; see exercise 1d.

In the solution of these normal equations by the routine exhibited in section 16, the minimized $\sum w(y_{obs} - y_{calc})^2$ appears in the left-most entry of Eq. III, as usual. Directly above it in Eq. 5 comes the reduction brought about by changing a from 1 to its final value, and in Eq. 6 appears the further reduction accomplished by turning the logarithmic line from the horizontal through the angle arc tan B.

Exercise 16. In fitting the equation

$$y = ae^{bx}$$

with x and y both subject to error, we may take F as in exercise 13, whereupon

$$F_o \equiv Y - a_o e^{b_o X} \qquad \text{(X and Y observed)}$$

Here we have use for the additional derivative $F_x = -by$, whence

$$L \text{ or } 1/W = b^2 y^2/w_x + 1/w_y$$

(If x is free of error, the 1st term of 1/W drops out and leaves $W = w_y$, the situation assumed for exercise 14; if y is free of error, the 2d term drops out and leaves $W = w_x/b^2y^2$).

Since we are here taking the case where x and y may both be in error, we set up table 1 of section 16 with headings as follows:

h	$F_x=-by$	w_x	w_y	L or 1/W	$\sqrt{L}$ or $1/\sqrt{W}$	$F_a = -y/a$	$F_b = -xy$	F_o

From this is formed table 2 with headings exactly like those of exercise 14 but with w replaced by W. Likewise, the normal equations will be symbolized as in exercise 14, w replaced by W. In fact, once table 2 is set up, from then on it is immaterial to the computer whether one or both coordinates are subject to error--a statement that holds good in any problem of curve fitting.

The solution of the normal equations will give v_a and v_b. The minimized sum of weighted squares (ϕ^2) will appear in the left-most entry of Eq. III, the portions removed by the successive adjustments of a and b falling in Eqs. 5 and 6 directly above ϕ^2; cf. exercise 3 of

section 17 and exercises 1, 4, 5, 10-15 of this section. Here, $\phi^2 = \Sigma(w_x V_x^2 + w_y V_y^2)$, both x and y residuals being present.

Exercise 17. To use the logarithmic form

$$\log y = \log a + Bx \qquad (B = b \log e = 0.434\ b)$$

or

$$y' = a' + Bx$$

for fitting the exponential $y = ae^{bx}$ when x and y are both subject to error, one would define f as in exercise 15, whereupon

$$f_o = Y' - (a_o' + B_o X)$$

The derivatives of f are as in exercise 15. L or 1/W will now have two terms, both coordinates being subject to error; in fact

$$L \text{ or } 1/W = B^2/w_x + 1/w_y'$$

w_y' being the weight of log y. The normal equations will be symbolized exactly like those of exercise 15a, but with W in place of w_y'. The left-most entry in Eq. III will be the minimized sum of squares ϕ^2, with the remarks at the end of exercise 16 applying here as well.

The analogy with exercise 4 is perfect throughout.

Exercise 4	Exercise 17
$y = a + bx$	$y' = a' + Bx$
$1/W = b^2/w_x + 1/w_y$	$1/W = B^2/w_x + 1/w_y'$

All the remarks and notes of exercise 4 apply here if y is replaced by y' or log y, a by a', b by B, and w_y by w_y'.

It is possible that in some problems, w_y' might be constant from one point to another (in which case w_y is not constant); then if w_x is also constant, we have a situation toward which the remark at the end of exercise 8 is directed.

Exercise 18. (a) Take

$$f = \log y - (\log a + Bx) \qquad \text{as in exercise 15}$$

$$F = y - ae^{bx} \qquad \text{" " " 13}$$

and suppose that x, y, a, and b take on small increments denoted by δx, etc. Prove that

$$\delta F = y\, \delta f \log e$$

hence at any point, $F_o = yf_o \log e$ to within higher powers of f_o of F_o.

(b) Thence prove that the normal equations in exercise 14 for fitting $y = ae^{bx}$ will give the same curve, i. e., the same results for a and b and for $\Sigma w(y_{obs} - y_{calc})^2$, as the normal equations in exercise 15a for the equivalent logarithmic form (exercise 15), except for discrepancies involving the squares and higher powers of residuals, the logarithmic form being slightly more accurate.* (Hint: note that if v_a is small, $v_{a'} = 0.434\, v_a/a$. Also, so far as a and b are concerned, the top normal equation in exercise 14 may be multiplied through by a).

The same comparison holds between exercises 16 and 17; but with $b \neq o$, and x and y both subject to error, there is not so much advantage in the logarithmic form.

Remark It may be worth while to pause for a thought on the factor $(2.30\ y)^2$ which is required for the proper weighting of log y (see exercise 15). Consider the following three points, artificially constructed for the purpose:**

x	y	log x	log y
1	10	0	1
10	7	1	0.7782
100	10	2	1
Av.	= 9	Av.	= 0.9261

x free of error, all three values of y have weight unity.

---

* My former pupil Mr. K. A. Norton first pointed this out to me.

**This illustration was developed in some correspondence with Professor W. L. Gaines of the University of Illinois, extending between 1932 and 1938.

Now it is perhaps evident that in fitting

$$y = ae^{bx}$$

or $$\log y = \log a + Bx \qquad (B = b \log e)$$

the result must be B = 0, because there are only three points, the end values of y being equal. One might therefore be tempted to take the data and fit the horizontal line

$$y = a \qquad (\text{since } e^{\circ} = 1)$$

getting $$a = \bar{y} = 9$$

with $$\sum res^2 = (10-9)^2 + (7-9)^2 + (10-9)^2$$

$$= 10^2 + 7^2 + 10^2 - 3 \cdot 9^2 = 6$$

This result is in fact correct.

It is interesting to see the same result come from the logarithms, correctly weighted. With B = 0, the equation for log a (from the top normal equation in the note appended to exercise 15, or simply by weighting log y proportional to y^2) is

$$[y^2] \log a = [y^2 \log y]$$

That is, $$(10^2 + 7^2 + 10^2) \log a = (10^2 + 7^2 \times 0.7782 + 10^2)$$

whence $$\log a = 238.14/249 = 0.95635$$

and $$a = 9.044$$

which is a trifle high, but good. The small discrepancy with the exact result (a = 9) arises from the size of the residuals in y, which are hardly small, though neither huge. Smaller residuals would have given closer agreement; larger ones would have produced a greater discrepancy. The factor y^2 has a tendency to over-weight the large values of y when the corresponding residual is large, but this overcompensation is greatly to be preferred over none at all, as will be evident in the next paragraph, wherein the weighting factor is ignored.

If one forgot to weight his logarithms in proportion to y^2, he should have had

$$\log a = [\log y]/3 = (1 + 0.7782 + 1)/3 = 0.9261$$

and $$a = 8.435$$

which is spurious, being considerably below the correct value a = 9.

The factor y^2 takes care of the change in scale that accompanies the taking of logarithms. The y values may all have the same weight, but their logarithms do not. No matter in what form the relation between x and y be written, the two terms F_xF_x/w_x and F_yF_y/w_y in L or 1/W can be relied upon to perform the same service as $(2.30\ y)^2$ does for the logarithmic scale.

This is an illustration of the fact that if the procedure of section 16 be followed, it makes no difference how a formula is written; one form will give the same curve as another form--except for disturbances arising from the neglect of 2d and higher powers of the residuals, but these are not usually of much consequence.

THE EXPONENTIAL WITH A LINEAR COMPONENT

Exercise 19. Given the equation

$$y = ae^{bx} + cx + d$$

to be fitted to n points. Write

$$F \equiv y - ae^{bx} - cx - d$$

The derivatives are

$$F_a = -e^{bx}, \quad F_b = -x\ ae^{bx}, \quad F_c = -x, \quad F_d = -1$$

$$F_x = -abe^{bx} - c, \quad F_y = 1$$

$$1/W = (abe^{bx} + c)^2/w_x + 1/w_y$$

(The 1st term of 1/W is missing if x is free of error, the 2d if y is free of error).

The formation of tables 1 and 2 of section 16, and the formation of the normal equations and their solution, is by this time a well-learned routine process. Here there are four parameters and hence unknowns, v_a, v_b, v_c, v_d. The derivatives F_a, F_b, etc., also F_o, for tables 1 and 2, are calculated with the approximate values a_o, b_o, c_o, and d_o, arrived at somehow, as by graphical methods, or previous experience.

The left-most entry in Eq. V of the solution of the normal equations will give the minimized $\Sigma(w_x V_x^2 + w_y V_y^2)$. The entries just above it in Eqs. 12, 13, 14, and 15 will show the reductions in the weighted sum of squares arising from the successive adjustments of a, b, c, and d.

Note that if only the y coordinates are subject to error, the left-most entry in Eq. V will give $\Sigma w \cdot res^2$, the deviations being measured in the vertical (i. e. parallel to the y axis). Moreover, if all y coordinates have the same weight (unity), then W = 1 throughout, and tables 1 and 2 of section 16 coalesce.

Note that with a formula of this kind, there is no possibility of making it linear by such a device as taking logarithms, and for this reason, it and others like it have been called insoluble. Fortunately the solution is entirely straightforward.

THE GENERALIZED HYPERBOLA AND ITS LOGARITHMIC FORM

Exercise 20 Given the equation

$$y = ax^b$$

to be fitted to n points. Here we write

$$F \equiv y - ax^b$$

= y obs

whence

$$F_o = y_{obs} - ax_{obs}^b \qquad \text{(a and b being replaced by } a_o \text{ and } b_o\text{)}$$

The derivatives of F are

$$F_x = -by/x, \qquad F_y = 1,$$

$$F_a = -y/a, \qquad F_b = -y \ln x.$$

$$L \text{ or } 1/W = b^2 y^2/a^2 w_x + 1/w_y$$

The headings for table 1 of section 16 in the general case would be

h	$F_x = -by/x$	w_x	w_y	L or 1/W	$\sqrt{L}$ or $1/\sqrt{W}$

(cont'd.)

$F_a = -y/a$	$F_b = -y \ln x$	F_o

It is easy to make the necessary modifications for special cases. Thus, if x is free of error, then $W = w_y$ and the F_x and w_x columns are superfluous; if further, all y coordinates have equal weight (unity) then $W = 1$ for all points, and tables 1 and 2 will coalesce. On the other hand, if y is free of error, then $1/W = b^2y^2/a^2w_x$ and the w_y column is omitted. From table 1 is formed table 2 with headings

h	$\sqrt{W}F_a$ or $F_a/\sqrt{L}$	$\sqrt{W}F_b$ or $F_b/\sqrt{L}$	$\sqrt{W}F_o$ or $F_o/\sqrt{L}$	Sum

The usual sums of squares and cross-multiplications from table 2 give the normal equations

No.	v_a	v_b	= 1	C_1	C_2	Sum
1	$[WF_aF_a]$	$[WF_aF_b]$	$-[WF_aF_o]$	1	0	-
2		$[WF_bF_b]$	$-[WF_bF_o]$	0	1	-
3			$[WF_oF_o]$	0	0	-

<u>Exercise 21.</u> The equation $y = ax^b$ of the preceding exercise may be turned into the logarithmic form

$$\log y = \log a + b \log x$$

or

$$y' = a' + bx' \qquad \text{(as in Ex. 15)}$$

Let

$$f = y' - (a' + bx')$$

$$f_o \text{ as usual}$$

The derivatives of f are

$$f_x = -0.434\, b/x \qquad f_y = 0.434/y$$

$$f_{a'} = -1, \qquad f_b = -x'$$

Then

$$L \text{ or } 1/W = 0.434^2\,(b^2/x^2w_x + 1/y^2w_y)$$

$$= b^2/w_x' + 1/w_y'$$

wherein $1/w_x' = 0.434^2/x^2 w_x$ = 1/wt. of x' or log x (see exercise 8c on page 30)

$1/w_y'$ similarly defined.

The headings for table 1 of section 16 will be

h, or Point No.	w_x	w_x'	w_y	w_y'	L or 1/W	$\sqrt{L}$	f_x	f_b	f_o

(See the remarks under table 1 of the preceding exercise. $f_{a'}$ is not listed, being constant).

The headings of table 2 will be the usual ones, as in table 2 of the preceding exercise with f in place of F.

The normal equations will be symbolized precisely like those of Set 1 in exercise 4. In fact all the remarks and notes of exercise 4 can be translated directly to the present problem. The reason is obvious: we have here a line in the variables x' and y', with weights w_x' and w_y'. The two terms of L or 1/W seen above take care of the change in the form of the function from exponential to logarithmic. In fact, we could say that

$$1/W = f_{x'}f_{x'}/w_{x'} + f_{y'}f_{y'}/w_{y'} = f_x f_x/w_x + f_y f_y/w_y$$

as in exercise 11b page 30.

Exercise 22. Prove that the normal equations of exercise 20 for fitting $y = ax^b$ will give the same curve, i. e. the same results for a and b and hence for ϕ^2, as the normal equations of exercise 21 for the equivalent logarithmic form, log y = log a + b log x, except for discrepancies involving the squares and higher powers of residuals, the logarithmic form (exercise 21) being slightly more accurate, especially if x is free of error. (Refer back to exercise 18).

THE HYPERBOLA WITH A LINEAR COMPONENT

Exercise 23. Given the equation

$$y = ax^b + c + dx$$

to be fitted to n points. Write

$$F \equiv y - ax^b - c - dx$$

F_o at any point is found, as usual, by giving x and y their observed values at that point, and a, b, c, d their approximate values a_o, b_o, c_o, d_o (found somehow).

The derivatives of F are

$$F_x = -abx^{b-1} - d, \quad F_y = 1,$$

$$F_a = -x^b, \quad F_b = -ax^b \ln x, \quad F_c = -1,$$

$$F_d = -x$$

whence

$$L \text{ or } 1/W = (abx^{b-1} + d)^2/w_x + 1/w_y$$

$$(W = w_y \text{ if x is free of error})$$

Tables 1 and 2 of section 16 are made up, and the normal equations formed and solved by the usual routine. Eq. V in the "1" column will give the minimized value of ϕ^2 or $\Sigma(w_x V_x^2 + w_y V_y^2)$, the successive reductions in the weighted sum of squares appearing in Eqs. 12, 13, 14, and 15, as usual (see exercises 1, 4, 5, 10-16).

The reader should refer back to the notes appended to exercise 19, which apply here as well.

Exercise 24. Given the equation

$$u = ax + by^c + dz$$

u, x, y, and z possibly all being observed. (This equation is used by Professor W. L. Gaines at the University of Illinois in his work on nutrition and lactation). Take

$$F = u - (ax + by^c + dz)$$

F_o as usual

The derivatives of F are

$$F_u = 1, \quad F_x = -a, \quad F_y = -c\ by^{c-1}, \quad F_z = -d$$

$$F_a = -x, \quad F_b = -y^c, \quad F_c = -by^c \ln y, \quad F_d = -z$$

$$\frac{1}{W} = \frac{1}{w_u} + \frac{a^2}{w_x} + \frac{(cby^{c-1})^2}{w_y} + \frac{d^2}{w_z}$$

Here we have a problem in four dimensions; 1/W contains four terms. The first term is absent if u is free of error, the second if x is free of error, etc.

The headings for table 1 in section 16 would be as follows:

h, or Point No.	w_u	w_x	by^c	F_y	w_y	w_z	$1/W$

cont'd.

$1/\sqrt{W}$	$F_a = -x$	$F_b = -y^c$	$F_c = -by^c \ln y$	$F_d = z$	F_o

Some of these headings will be omitted if any of the u, x, y, or z values are free of error throughout.

For table 2, for by obvious divisions from table 1:

h, or Point No.	$\sqrt{W}F_a$	$\sqrt{W}F_b$	$\sqrt{W}F_c$	$\sqrt{W}F_d$	$\sqrt{W}F_o$	Sum

The normal equations will be formed and solved in the usual routine manner. Eq. V will contain the minimized value of ϕ^2 or $\Sigma(w_uV_u^2 + w_xV_x^2 + w_yV_y^2 + w_zV_z^2)$ in the "l" column, the successive

reductions in the weighted sum of squares appearing in Eqs. 12, 13, 14, and 15, as usual; see exercises 1, 4, 5, 10-16, 23).

If only the u coordinates are subject to error, $\phi^2 = \sum w_u V_u^2$. Moreover, if all the u coordinates have the same weight (unity), then W = 1 throughout, and tables 1 and 2 coalesce.

The last note appended to exercise 19 applies here.

MISCELLANEOUS

Exercise 25. (a) Given the equation

$$u = ax + by + cz$$

to be fitted to n observed points. With

$$F = u - (ax + by + cz)$$

show that

$$1/W = 1/w_u + a^2/w_x + b^2/w_y + c^2/w_z$$

and that the normal equations are

No.	v_a	v_b	v_c	=	1	C_1	C_2	C_3	Sum
I	[Wxx]	[Wxy]	[Wxz]		$-[WxF_o]$	1	0	0	-
2		[Wyy]	[Wyz]		$-[WyF_o]$	0	1	0	-
3			[Wzz]		$-[WzF_o]$	0	0	1	-
4					$[WF_oF_o]$	0	0	0	-

(b) If it is desired to solve for a, b, c directly, without the use of the approximations a_o, b_o, c_o, the unknowns in the normal equations would be a, b, and c, and the "1" column would be

[Wxu]
[Wyu]
[Wzu]
[Wuu]

extra decimals being required for accuracy, as mentioned on page 119.

(c) Prove that the minimized sum of weighted squares is

$$\phi^2 = [Wuu] - [Wxu]a - [Wyu]b - [Wzu]c$$

(See exercise 3a in section 17)

<u>Remark.</u> If u alone is subject to error, and of uniform precision (unit weight) throughout, the only change in the normal equations would be that W would not appear, being unity throughout. The minimized sum of squares would be

$$\phi^2 = [uu] - [xu]a - [yu]b - [zu]c$$

(This equation is used a good deal in some kinds of statistical work; see e. g. page 160 of the 6th edition of Fisher's <u>Statistical Methods for Research Workers</u>, on which the above equation appears as

$$S(y-Y)^2 = S(y^2) - b_1 S(x_1 y) - b_2 S(x_2 y) - b_3 S(x_3 y)$$

this being the sum of squares after fitting

$$Y = b_1 x_1 + b_2 x_2 + b_3 x_3)$$

See example 3 in section 20 for an illustration.

20. Three examples in curve fitting.

Example 1, fitting an isotherm. Data taken by Michels et al., Proc. Royal Society (London) 153A, 201-224, 1935. The equation to be fitted is

$$y = a + bx + cx^2 + dx^4$$ (y denotes pv, pressure times volume; x denotes density)

The parameters are not independent but are subject to the condition that

$$y = 1 \text{ when } x = 1$$

Then $$a = 1 - b - c - d$$

and $$y = 1 + (x-1)b + (x^2-1)c + (x^4-1)d$$

Weights: All y coordinates to have equal weight; x free of error.

Let $$F = y - \{1 + (x-1)b + (x^2-1)c + (x^4-1)d\}$$

Derivatives: $F_b = -x + 1, \quad F_c = -x^2 + 1, \quad F_d = -x^4 + 1$

$F_y = 1$, $\quad F_x$ is not needed since x is free of error.

$W = 1$ at every point.

The following approximate values are known from previous experience:

$$b_o = -0.006837046$$

$$c_o = 0.000011392$$

$$d_o = 0.0^9 1514$$

Let $$y' = 1 - .006837046\,(x-1) + .000011392(x^2-1) + .0^9 1514(x^4-1)$$

Then $$F_o = y_{obs} - y'$$

Tables 1 and 2* (formed from the original data).

Point	$-F_b$	$-F_c$	$-F_d$	$-F_o$	Sum
	1.77x10	3.51x10^2	1.24x10^5	0.55x10^{-4}	
1	1.77	3.51	1.24	0.55	7.07
2	2.25	5.49	3.03	0.79	11.56
3	2.72	7.93	6.30	2.07	19.02
4	3.18	10.74	11.55	3.09	28.56
5	3.66	14.15	20.04	6.22	44.07
6	4.14	17.99	32.39	10.05	64.57
7	4.61	22.20	49.31	14.84	90.96
Sum	22.33	82.01	123.86	37.61	265.81✓

The normal equations are formed in the usual manner by cumulating squares and cross-multiplications from the above table. The powers of 10 in the headings bring uniformity in the denominations of the columns. The Sum column provides a check, which should never be omitted; it is formed regardless of the powers of 10; in fact no attention is paid to the powers of 10 until the end, when the solution is decoded. Why is it better to start off with 100 rather than 1 in the C_1, C_2, and C_3 columns of Eqs. I, 2, and 3, for the calculation of the reciprocal matrix? Perhaps 1000 would have been better than 100.

Note the symmetry in the reciprocal matrix, which is found between the vertical lines in Eqs. 11, 12, and 13.

---

* In this example, W = 1 throughout, with the result that tables 1 and 2 mentioned in section 16 are identical. The minus signs in the headings avoid minus signs in the table.

Normal equations: (the solution follows the form outlined in section 19).

	$10v_b$	$10^2 v_c$	$10^5 v_d$	= 1	C_1	C_2	C_3	Sum
I	77.53	302.94	497.86	$151.06\cdot10^{-4}$	100	0	0	1129.38✓
2		1236.98	2155.67	654.02	0	100	0	4449.61✓
3			4066.38	1233.79	0	0	100	8053.70✓
4				374.68	0	0	0	2413.55✓
	Factors							
5	-3.907377	-1183.68	-1945.32	-590.23	-390.74	0	0	-4412.90
II		53.30	210.35	63.79	-390.74	100	0	36.70✓
		-1	[illegible]	-1.197				
6	-6.421556		-3197.02	-970.01	-642.16	0	0	-7252.36
7	-3.946701		-830.20	-251.75	1542.12	-394.67	0	-144.85
III			39.16	12.03	899.97	-394.67	100	656.49✓
			-1	-.307				
8	-1.948379			-294.31	-194.84	0	0	-2200.46
9	-1.196797			-76.34	467.63	-119.68	0	-43.93
10	-0.3071320			-3.69	-276.41	121.22	-30.71	-201.63
IV				.33	-3.61	1.54	-30.71	-32.46✓

Therefore $\Sigma(wV^2) = .33\text{x}10^{-8}$; $\sigma^2(\text{ext}) = .33\text{x}10^{-8}/4 = .082\text{x}10^{-8}$

13	$10v_b$ =			.036136	236.75	-98.03	22.98	
12		$10^2 v_c$ =		-.015361	~~-98.03~~	41.65	-10.08	
11			$10^5 v_d$	.3071320	~~22.98~~	~~-10.08~~	2.55	16.76✓

$\therefore\ v_b = +\ .036\text{x}10^{-5}$ $\qquad b = -.00683705\ -\ .0^{6}36\ =\ -.00683741$

$v_c = -\ .015\text{x}10^{-6}$ $\qquad c = .0^{4}11392\ +\ .0^{7}15\ =\ .0^{4}11407$

$v_d = +\ .307\text{x}10^{-9}$ $\qquad d = .0^{9}151\ -\ .0^{9}307 = -.0^{9}156$

$\therefore\ a = 1 - b - c - d = 1.00682600$

Est'd S.E.^{2} of $b = 236.75 \text{ x } 10^{-2-2}\ \sigma^2(\text{ext}) = 19.3 \text{ x } 10^{-12}$

" " " $c = 41.65 \text{ x } 10^{-2-4}\ \sigma^2(\text{ext}) = 3.40 \text{ x } 10^{-14}$

" " " $d = 2.55 \text{ x } 10^{-2-10}\sigma^2(\text{ext}) = 0.209\text{x } 10^{-20}$

" " " $a = \sigma^2(\text{ext})(236.75 \text{ x } 10^{-4} + 41.65 \text{ x } 10^{-6} + 2.55 \text{ x } 10^{-12}$

$-2 \text{ x } 98.03 \text{ x } 10^{-5} + 2 \text{ x } 22.98 \text{ x } 10^{-8} - 2 \text{ x } 10.08 \text{ x } 10^{-9})$

$= 17.7 \text{ x } 10^{-12}$ (see Eq. 101; remember that a is a function of b, c, and d. See also exercise 1 following).

The final results are:

$a = 1.0068260 \pm .0000042$
$b = -.0068374 \pm .0000044$
$c = .0^{4}1141 \pm .0^{6}18$
$d = -.0^{9}156 \pm .0^{9}046$

S.Es. estimated from 4 degrees of freedom.

Exercise 1. Show that when corrected for powers of 10 the reciprocal matrix is

$$A^{-1} = \begin{vmatrix} 236.75 \times 10^{-4} & -98.03 \times 10^{-5} & 22.98 \times 10^{-8} \\ -98.03 \times 10^{-5} & 41.65 \times 10^{-6} & -10.08 \times 10^{-9} \\ 22.98 \times 10^{-8} & -10.08 \times 10^{-9} & 2.55 \times 10^{-12} \end{vmatrix}$$

These are the figures that were used in writing down the S. Es. of a, b, c, and d. Evaluated as a determinant, this gives $A^{-1} = 4.6 \times 10^{-22}$.

Exercise 2. The evaluation of the determinant of the coefficients is

$$A = (77.53 \times 10^{2})(53.30 \times 10^{4})(39.16 \times 10^{10}) = 0.162 \times 10^{22}$$

(See Ex. 1 of section 17, page 108). This result is not exact; the discrepancy arises from instability, and could be overcome by carrying more decimals.

Exercise 3. Prove the S.E. of the curve at x = 1 is 0, and that at x = 0 it is the same as the S.E. of a. Why is the S.E. of the y intercept practically equal to the S.E. of b? Argue geometrically.

As often happens in curve fitting, these normal equations are unstable. The trouble comes from the fact that there is not enough spread in the abscissas to determine accurately the coefficient d. One of the most sensitive tests for instability is to compare the direct solution (already found in the "1" column of Eqs. 11, 12, and 13) with that given by using the reciprocal matrix as a multiplier; by such means we get the reciprocal solution (pp. 111-112)

$$10v_b = \{(151.06)(236.75) - (654.02)(98.03) + (1233.8)(22.98)\}\ 10^{-2-2}$$
$$= 0.0237 \cdot 10^{-4}$$

and in like manner (which the student should undertake as an exercise),

$$10^2 v_c = -0.0508 \cdot 10^{-4}, \qquad 10^5 v_d = 0.2500 \cdot 10^{-4}$$

These are in disagreement with the direct solution found in Eqs. 11, 12, and 13, and thus instability is indicated. The direct solution satisfies the normal equations to the last decimal, but when there is instability, many other solutions not too far away could do the same thing.* The reciprocal solution, however, does not satisfy the normal equations, the actual numbers being 111 against 151.06, 483 against 654.02, 919 against 1233.79.

The insidious thing about instability is that its presence may go undetected. For instance, if here we had only the "reciprocal solution", and had not tried to check it by substitution, we might have accepted it. The use of the reciprocal matrix as a multiplier is in theory very fascinating, but as a practical matter in curve fitting we should not lose sight of the fact that it occasionally does more harm than good. Fortunately it does work to good advantage in many problems, as seen for instance in Ch. 5 of R. A. Fisher's <u>Statistical Methods for Research Workers</u>. In section 12d, also, the equations were stable and no difficulties arose. It is interesting to see what would be the sum of squares if the term dx^4 had been dropped. From Eq. III we find [cc.2] = 39.16×10^9; this multiplied by d_0^2 or $(0.0^9 1514)^2$ gives 0.09×10^{-8}, which added to 0.33 gives 0.42×10^{-8} for the sum of the $(y_{obs} - y_{calc})^2$ that would have resulted from fitting the curve $y = a + bx + cx^2$. We then find

$$\sigma^2(\text{ext}) = 0.42 \times 10^{-8}/5 = 0.084 \times 10^{-8} \quad (dx^4 \text{ omitted})$$

and we already had

$$\sigma^2(\text{ext}) = 0.082 \times 10^{-8} \qquad\qquad (dx^4 \text{ included})$$

So, as we might expect from the instability of the equations, the indications are that the data are insufficient to determine d very accurately, in spite of the fact that the value of d turns out to be triple its estimated S.E. Other solutions could be found for the normal equations, satisfying them to the last decimal, but allowing d to be considerably different from that which we happened to find here.

---

* See the author's paper in Science, 7th May 1937, cited in section 17. Further work by L. B. Tuckerman of the National Bureau of Standards is in progress and may appear in Science in 1938 or 1939.

Example 2 (illustration of Ex. 10 of section 19). The polynomial

$$y = a + bx + cx^2$$

fitted to the following points in the x, y plane, x and y both subject to error.

Observations

X and Y denote the averages of N readings. s = S.D. of N readings

Point No.	On the x coordinates X(cm.)	N	s^2	On the y coordinates Y(lb.)	N	s^2
1	-1.92	5	0.306	0.087	10	.00258
2	-1.15	6	.336	.141	9	.00320
3	.04	7	.117	.225	8	.00264
4	.86	8	.232	.234	7	.00313
5	2.01	9	.265	.267	6	.00126
6	2.99	10	.173	.349	5	.00167
7	3.69	9	.186	.435	6	.00064
8	5.20	5	.428	.460	10	.00205
9	6.09	7	.096	.549	8	.00209
10	7.18	6	.098	.580	9	.00283
11	7.80	5	.192	.638	10	.00152
12	9.08	6	.179	.728	9	.00078

The table shows, for instance, that at point No. 1, there were 5 observations on the x coordinate, their average being X = -1.92, and their S.D.^{2} = 0.306.

The question arises how to weight the various values of X and Y. For one thing, the weight of any coordinate will be proportional to N, but that is not enough; the precision of single observations may vary from point to point, and it is evidently different for the y coordinates than for the x coordinates, judging from the s^2 columns.

Previous experience in this line of work, let us suppose, has indicated that the S. E. of single observations on x is about 0.5 cm., and on y about 0.05 lb. In order to investigate the precisions for this particular set of observations, we make a graph of the values of $s^2N/(N-1)$ for x, and of the same thing for y, both plotted against x (y would do as well). $s^2N/(N-1)$ for x (or y) at any point is an estimate of the square of the S. E. of the single observations on x (or y) at that point.

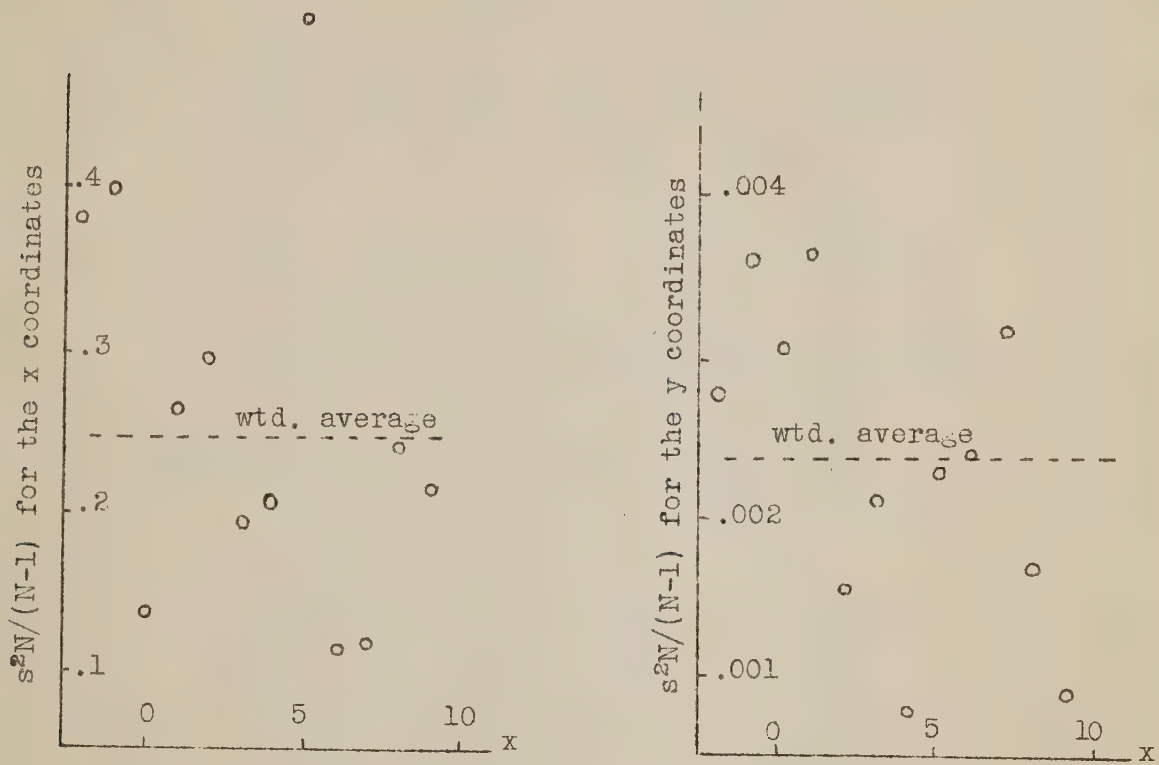

Fig. 12. Estimating the precision of the observations

Although there is fluctuation of the estimates, there is no trend; moreover, the weighted average on the x plot is not far from 0.25, and the average on the y plot is not far from 0.0025, so it seems reasonable to conclude that the prior values of precision should not be changed. Accordingly we take 0.5 for the S. E. of single observations on x, and 0.05 for the S. E. of single observations on y, over the entire range. By recalling that weights are inversely proportional to the squares of variances, we see that a single observation on a y coordinate has 100 times the weight of a single observation on an x coordinate (v. Eq. 16 p. 13). As a matter of convenience, then, we take unity for the weight of a single observation on x, and 100 for the weight of a single one on y. This is equivalent to setting $\sigma = 0.5$ for observations of unit weight. The values of w_x (next page) are then the same as the numbers N on the preceding page referring to x, and the values of w_y are 100 times the numbers N referring to y. (If the precision of single observations on either x or y coordinates were variable over the range of the points, obvious modifications could be made in the weighting).

For the function being fitted we take

$$F = y - (a + bx + cx^2)$$

whence $F_x = -(b+2cx)$, $F_y = 1$

$F_a = -1$, $F_b = -x$, $F_c = -x^2$ (See exercise 10 of section 19)

and $1/W = (b+2cx)^2/w_x + 1/w_y$

By passing the curve $y = a + bx + cx^2$ through points numbered 1, 6, and 11, the approximate values

$$a_o = 0.1855$$

$$b_o = 0.0526$$

$$c_o = 0.00069$$

are found (see note on the "method of selected points" page 34). This selection makes $F_o = 0$ at points 1, 6, and 11 in table 1.

We are now ready to fill in table 1, the 3d step of section 16. $1/W$ is computed by using the value of $(b+2cx)^2$ in the 2d column (see the formula for $1/W$ written out above).

Table 1.

Point No.	$(b+2cx)^2$ **	w_x	w_y	$1/W$ **	$1/\sqrt{W}$	$-F_a$ *	$-F_b$	$-F_c$	F_o **
1	0.00253	5	1000	0.00151	0.039	1	-1.92	3.69	0
2	264	6	900	155	39	1	-1.15	1.32	+0.015
3	281	7	800	165	41	1	0.04	0	+ .037
4	294	8	700	180	42	1	0.86	0.74	+ .003
5	312	9	600	201	45	1	2.01	4.04	- .027
6	327	10	500	233	48	1	2.99	8.94	0
7	338	9	600	204	45	1	3.69	13.62	+ .046
8	363	5	1000	173	42	1	5.20	27.04	- .018
9	379	7	800	179	42	1	6.09	37.09	+ .017
10	398	6	900	177	42	1	7.18	51.55	- .019
11	409	5	1000	182	43	1	7.80	60.84	0
12	432	6	900	183	43	1	9.08	82.45	+ .008

* Since $-F_a = 1$ all the way down, it would ordinarily not be listed.

** Computed with $a_o = 0.1855$, $b_o = 0.0526$, $c_o = 0.00069$. Why is $F_o = 0$ at points No. 1, 6, and 11?

The 4th step of section 16 is to form

Table 2 (The matrix*)

Point No.	$-\sqrt{W}\,F_a$	$-\sqrt{W}\,F_b$	$-\sqrt{W}\,F_c$	$\sqrt{W}\,F_o$	Sum
1	25.8	-4.9 x 10	0.9 x 10^2	0	21.8
2	25.4	-2.9 "	0.3 "	3.8 x 10^{-1}	26.6
3	24.6	0.1 "	0	9.1 "	33.8
4	23.6	2.0 "	0.2 "	0.7 "	26.5
5	22.3	4.5 "	0.9 "	-6.0 "	21.7
6	20.7	6.2 "	1.8 "	0	28.7
7	22.1	8.2 "	3.0 "	10.2 "	43.5
8	24.0	12.5 "	6.5 "	-4.3 "	38.7
9	23.6	14.4 "	8.8 "	4.0 "	50.8
10	23.8	17.0 "	12.2 "	-4.5 "	48.5
11	23.5	18.3 "	14.3 "	0	56.1
12	23.4	21.2 "	19.3 "	1.9 "	65.8
Sum	282.8	96.6	68.2	14.9	462.5✓

Note that the minus signs at the headings obviate the writing of minus signs all the way down. Also note that the Sum" column is formed without regard to the powers of 10.

The next step is the formation of the normal equations by the usual cumulation of squares and cross-products from table 2. The solution is carried out by the routine outlined on page 107 and used previously on pages 64 and 65, and in the preceding example.

---

* So-called because from it is formed the normal equations. Moreover, in matrix notation, the formation of the normal equations is the same as taking the product M'M, M being the matrix of table 2 of section 16, and M' its "transpose".

Normal Equations

No.	$-v_a$	$-10v_b$	$-100v_c$	= 1*	C_1	C_2	C_3	Sum**
I	6686	2230	1601	357×10^{-1}	100	0	0	10974✓
2		1599	1121	16	0	100	0	5066✓
3			859	16	0	0	100	3697✓
4				296	0	0	0	684✓
	Factors							
5	-.33353	-744	-534	-119	-33.35	0	0	-3660
II		855	587	-103	-33.35	100	0	1406✓
6	-.23946		-383	- 85	-23.95	0	0	-2628
7	-.68655		-403	71	22.90	-68.66	0	- 965
III			73	2	- 1.05	-68.66	100	104✓
8	-.053395			- 19	- 5	0	0	- 586
9	.12047			- 12	- 4	12	0	169
10	-.027397			0	0	2	-3	3
IV				265	- 9	14	-3	270✓
13	Subst. 11 & 12 into I			$-v_a$ = .09330	.02812	-.02915	-.01433	
12	Subst. 11 into II			$-10v_b$ = -.13930	-.02915	.7627	-.9406	
11	From Eq. III			$-100v_c$ = .02740	-.01438	-.9406	1.370	

From Eq. IV, $\phi^2 = 2.65$

From Eqs. 13, 12, and 11 we see that

$$v_a = -0.00933$$

$$v_b = 0.00139$$

$$v_c = -0.00003$$

The adjusted values of the parameters are therefore

$$a = 0.1855 + 0.0093 = 0.1948$$

$$b = 0.0526 - 0.0014 = 0.0512$$

$$c = 0.00069 + 0.00003 = 0.0007$$

---

* The factor 10^{-1} holds all the way down the "1" column.

** The decimals in the C_1, C_2, and C_3 columns were added later, and are ignored in the Sum column.

The matrix reciprocal to the coefficients in the normal equations is contained in Eqs. 11, 12, and 13 under C_1, C_2, and C_3; the powers of 10 must be adjusted, however, before it will give the variance and product variance coefficients for a, b, and c. Correcting for powers of 10 we have

$$A^{-1} = \begin{vmatrix} .0^3 28 & -.0^4 29 & -.0^5 14 \\ -.0^4 29 & .0^4 77 & -.0^5 95 \\ -.0^5 14 & -.0^5 95 & .0^5 14 \end{vmatrix}$$

for the variance and product variance coefficients of a, b, and c. The S.Es. of the parameters are then as follows ($\sigma = 0.5$):

$$(\text{S.E. of } a)^2 = 0.00028\,\sigma^2, \text{ or S.E. of } a = 0.0084$$

$$(\text{S.E. of } b)^2 = 0.000077\,\sigma^2, \text{ or S.E. of } b = 0.0044$$

$$(\text{S.E. of } c)^2 = 0.0000014\,\sigma^2, \text{ or S.E. of } c = 0.00059$$

The estimated parameters are then written

$$\left.\begin{aligned} a &= 0.1948 \pm 0.0084 \\ b &= 0.0512 \pm 0.0044 \\ c &= 0.0007 \pm 0.0006 \end{aligned}\right\} \text{S.Es. based on a prior value of } \sigma$$

On the basis of standard errors, it looks as if the value of c is barely significant. We may wonder if a straight line would fit about as well. If one were to cut off the term cx^2, the resulting increment in ϕ^2 would be roughly $[cc.2]c^2$, as we know from exercise 3 in section 17. The qualifying adjective roughly is needed here because the weighting factors W in table 1 contain c, and were calculated, not on the basis of c = 0, but with c = 0.00069. Seeing that [cc.2] in Eq. III is 73×10^{-4}, we accordingly add

$$73 \cdot 10^{-4} \cdot 0.0007^2 = 0.37$$

to 2.65 and obtain 3.02 for a rough value of what ϕ^2 would have been if we had fitted the line $y = a + bx$. The external estimate of σ would have been about

$$\sigma^2(\text{ext}) = 3.02/(12-2) = 0.30 \qquad \text{(roughly)}$$

whereas, for the fitted parabola $y = a + bx + cx^2$ it is

$$\sigma^2(\text{ext}) = 2.65/(12-3) = 0.29$$

The comparison of the two external estimates thus also suggests that so far as this set of data is concerned, the line is about as good a fit as the parabola. In other words, the term cx^2 is incapable of reducing the sum of squares by as much as it ought to if it were really needed. This situation causes some instability, as in example 1. The reciprocal solution is in fact $-v_a = -0.0934\cdot10^{-1}$, $-10v_b = -0.1325\cdot10^{-1}$, $-100v_c = 0.0175\cdot10^{-1}$, but we shall not stop to discuss it.

It will be recalled that here we were in the fortunate situation of having a reliable value of σ, both from previous experience and from the "internal consistency" of the observations.* The internal estimate was σ^2 (int) = about 0.25, this being the average ordinate in Fig. 12 on the basis that single observations on an x coordinate have weight unity, and single observations on a y coordinate have weight 10. So with σ^2(ext) = 0.29, the internal and external estimates are in good agreement, neither the function nor the data are laid open to any suspicion by a comparison of the two estimates of σ.

It is important to keep in mind the fact that when neither a prior value of σ nor an estimate by external consistency is at hand, σ(ext) is no help in judging the fit of the curve; one is bound to come out with $P(\chi)$ equal to about 0.5 if he computes χ^2 with σ replaced by σ(ext) whether the fit is good or bad, as already noted on page 19.** Nevertheless, two external estimates are useful by themselves for comparing two different formulas fitted to the same data--for instance, in deciding whether the term cx^2 is really needed (supra).

In this connection I should like to quote two paragraphs from a manuscript on the melting point of rhodium, written by my friends H. T. Wensel and L. B. Tuckerman of the National Bureau of Standards.

"It should be emphasized that although, lacking other knowledge, Gauss's criterion+ of closeness of fit is the best criterion we have, it is far from a certain criterion. It is based upon the assumption that the deviations of the observed points from both of the empirical curves are random samples drawn from a normal universe of errors. Unless the distribution of the residuals is a highly probable distribution for such a sample no confidence whatever can be placed even in this criterion.

---

* See reference to Birge in section 6c. A comparison of the internal and external estimates of σ is an application of the "analysis of variance", though historically it is interesting to note that physicists were using the notion long before the term "analysis of variance" was coined, note for instance the reference to Palmer on page 18.

** Possibly some remarks of mine in the Physical Review 49, 243-247, 1936, might be helpful.

\+ They refer to σ(ext).

"As a simple illustration, two non-intersecting parabolas indicating a putative discontinuity can be fitted to a subset of 8 and a subset of 6 values of the continuous function $y = \sin x$ chosen at equally spaced values of x, which on the basis of Gauss's criterion fit the points better than any single parabola fitted to all the 14 points of the set. The criterion, however, is in this case wholly invalid since the deviations of a sine curve from a parabola are not random but are systematic." (To appear in the J. Optical Society, 1938).

Now to adjust the observations, i.e. to estimate where the point is that was observed to be at X, Y, we note that there will be a λ at every point, given by Eqs. 87, page 86:

$$\lambda_1 = W_1(F_o^1 - F_a^1 v_a - F_b^1 v_b - F_c^1 v_c)$$

$$\lambda_2 = W_2(F_o^2 - F_a^2 v_a - F_b^2 v_b - F_c^2 v_c)$$

$$\vdots \qquad \vdots$$

(The superscripts on F refer to the point number)

We use $v_a = -0.00938$, $v_b = 0.0014$, $v_c = -0.00003$, and the F values from table 1, page 160. We shall not adjust all the observations, but only those at points No. 6, 7, and 8 for illustration. Using equations like the ones just written, we find that

$$\lambda_6 = \frac{1}{0.00233}(0 - 0.00938 + 2.99 \cdot 0.0014 - 8.9 \cdot 0.00003)$$

$$= -2.344$$

$$\lambda_7 = \frac{1}{0.00204}(0.046 - 0.00938 + 3.69 \cdot 0.0014 - 13.6 \cdot 0.00003)$$

$$= 22.32$$

$$\lambda_8 = \frac{1}{0.00173}(-0.018 - 0.00938 + 5.20 \cdot 0.0014 - 27.0 \cdot 0.00003)$$

$$= -10.91$$

whence by applying Eq. 89, p. 91, the residuals can be computed at once.

At No. 6: $V_x = (1/w_x)\lambda_6 F_x = (-2.344)(-.05549)/10 = 0.013$

$$V_y = (1/w_y)\lambda_6 F_y = -2.344/500 = -0.0047$$

At No. 7: $V_x = (1/w_x)\lambda_7 F_x = (22.32)(-.05647)/9 = -0.140$

$V_y = (1/w_y)\lambda_7 F_y = 22.32/600 = 0.0372$

At No. 8: $V_x = (1/w_x)\lambda_8 F_x = (-10.91)(-.0586)/5 = 0.128$

$V_y = (1/w_y)\lambda_8 F_y = -10.91/1000 = -0.0109$

These residuals measured off from the observed points in the proper direction (see Fig. 9, page 83) give the calculated points, which are supposed to lie on the calculated curve. Actually there are small discrepancies; the points so calculated do not fall exactly on the curve, as is barely evident in Fig. 13. But such discrepancies are trifling, being of second order from the neglect of squares and higher powers of the residuals. One may simply manipulate the end decimal of V_x or V_y or both, in order to place the calculated point

$$\left.\begin{aligned} x &= X - V_x \\ y &= Y - V_y \end{aligned}\right\} \qquad \text{[Eq. 90, p. 91]}$$

exactly on the curve. A precisely similar situation arises in problems of surveying, wherein for exact satisfaction of the geometrical conditions after adjustment, one often needs to manipulate the end decimal of one or more angles and sides (see page 68, middle).

The adjustment of the observations is important, and now that methods are available in the general case (Eqs. 89, page 91), it should always be carried out so that the residuals can be inspected before any conclusion is based on the summation of $w_x V_x^2 + w_y V_y^2$. As a matter of fact, in this particular problem, a plot of all 12 observed and calculated points directs no suspicion toward the randomness of the residuals, but this is always a matter of human judgment. If the term cx^2 is omitted, and a straight line fitted to the points, the graph of the residuals so obtained gives a faint suggestion that the parabolic term cx^2 should be included. In this particular line of investigation, previous experience has demanded the term cx^2, which is something this small set of data by itself could not settle.

Exercise. Compute the x and y residuals for the other 9 points, and plot them.

The S.E. of any function f(a, b, c) is

$$\sigma_f^2 = \sigma^2\{0.00028(\frac{df}{da})^2 + 0.000077(\frac{df}{db})^2 + 0.0000014(\frac{df}{dc})^2$$
$$- 2(0.000029)\frac{df}{da}\frac{df}{db} - 2(0.0000014)\frac{df}{da}\frac{df}{dc}$$
$$- 2(0.0000095)\frac{df}{db}\frac{df}{dc}\}$$

(see Eq. 47, p. 27)

In particular, the calculated y with its S.E. could be written

$$y = 0.1948 + 0.0512\ x + 0.0007\ x^2 \pm \sigma\{0.00028 + 0.000077\ x^2$$
$$+ 0.0000014\ x^4 - 0.000058\ x - 0.0000028\ x^2 - 0.000019\ x^3\}^{\frac{1}{2}}$$

(The terms $0.000077\ x^2$ and $-0.0000028\ x^2$ would of course be combined).

If the factor σ in front of the brace be replaced by $1.96\ \sigma$, the double sign gives a 95% "confidence" band (Neyman's terminology*), which interpreted means that at any abscissa, only 1 in 20 of the bands so calculated will fail to cover the true value of y, the assumption being that the observations actually are normally distributed with variances σ^2/w_x and σ^2/w_y about certain "true values". The 95% confidence band for this particular set of observations is shown dashed in Fig. 13.

---

* See pages 115 and 116, also J. Neyman, Washington Lectures, (cited on page 18), especially the conference on confidence intervals.

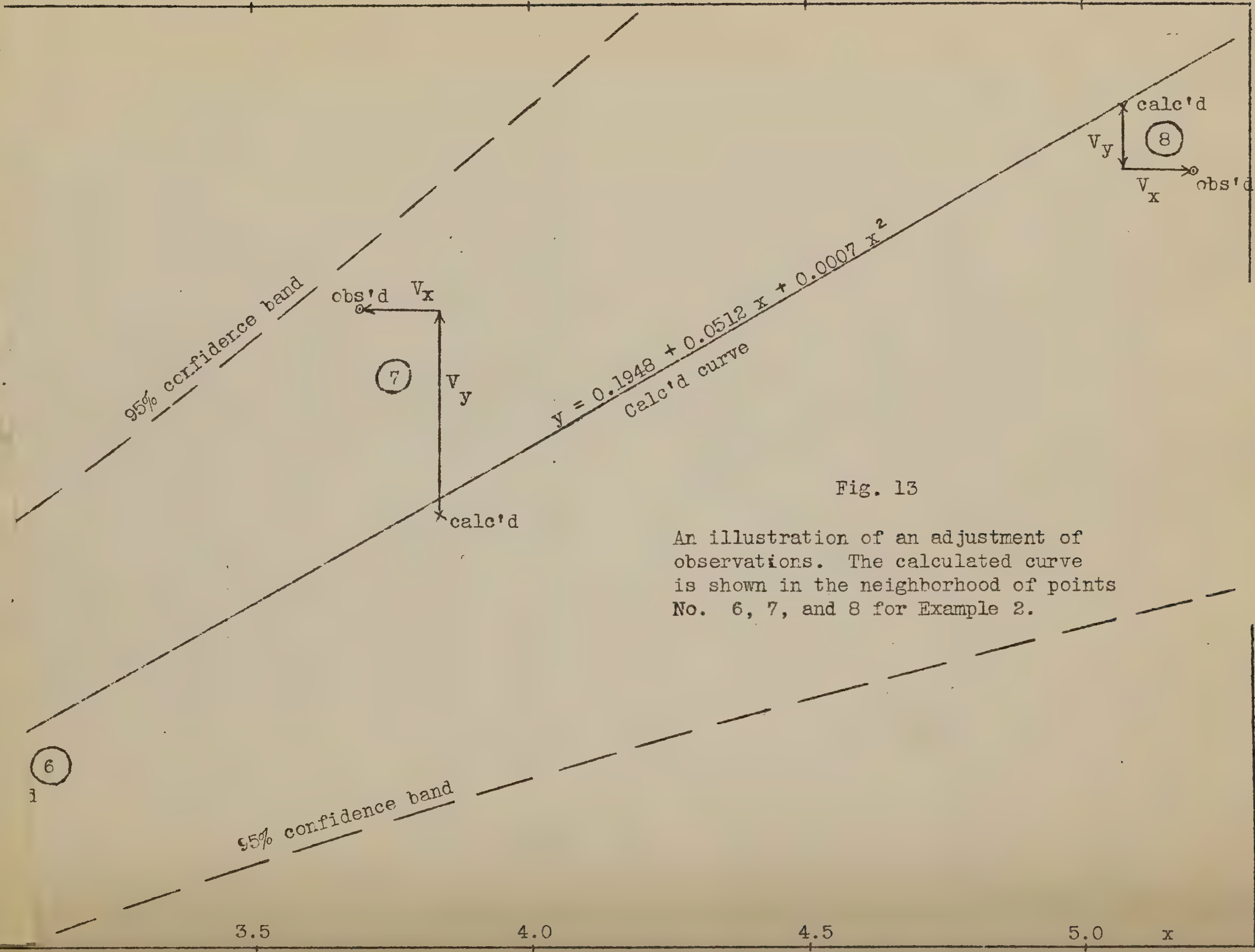

Fig. 13

An illustration of an adjustment of observations. The calculated curve is shown in the neighborhood of points No. 6, 7, and 8 for Example 2.

Example 3, a formula useful in forestry*.

This example serves the purpose of illustrating three features: (1°) the fitting can be done with logarithms, the constant or nearly constant characteristic being suppressed or nearly suppressed to cut down the number of figures required; (2°) W is constant throughout; (3°) the prior value of σ can be expressed in terms of some of the parameters, so that, finally, the minimized ϕ^2 can be transformed into χ^2, and the fit of the formula judged on this criterion. All three features owe their existence both to the form of the fitted function and to the experimental material and procedure. One or more of them, however, is apt to be encountered in other work.

If

x = the volume of a tree in board feet;

y = the merchantable height of the tree;

z = its diameter at breast height;

then experience has shown that the equation

$$x = ay^b z^c$$

expresses satisfactorily** the relation between x, y, and z.

The particular set of data for consideration in the present problem consists of 66 points -- measurements on the volume, merchantable height, and diameter, of 66 trees. It will not be necessary to display the full set of points for the discussion intended here; the first six and the last will be sufficient. They come in no particular order of size. The logarithms are written in the three right-hand columns, for convenient inspection, since they will be needed in the fitting.

Looking at the logarithms (top of the next page) we perceive that the first part of the figures, the "characteristics", do not vary much; in fact, in the Y' and Z' columns the characteristic is unity all the way down. What we need to do is to write the formula in such a way that the variable part of the logarithms is brought into prominence. This can be accomplished by writing the formula as

---

* I am indebted to my friend Mr. Jesse H. Buell of the Forest Service, Asheville, for bringing this problem to me.

** Francis X. Schumacher and F. dos S. Hall, J. Agric. Res. 47, 719-734, 1933; also Donald Bruce and Francis X. Schumacher, Forest Mensuration (McGraw-Hill, 1935) Art. 140.

	OBSERVATIONS			LOGARITHMS		
Point No.	Volume X (board feet)	Height Y (feet)	Diam. Z (inches)	X' = log X	Y' = log Y	Z' = log Z
1	60	25	13.8	1.778	1.398	1.140
2	60	24	14.0	1.778	1.380	1.146
3	120	29	18.1	2.079	1.462	1.258
4	270	38	21.0	2.431	1.580	1.322
5	320	37	21.6	2.505	1.568	1.334
6	130	30	16.5	2.114	1.477	1.218
:	:	:	:	:	:	:
:	:	:	:	:	:	:
66	320	54	18.8	2.505	1.732	1.274
			Sums	152.136	102.451	84.090

$$x' = a' + by' + cz' \qquad (x' = \log x, \text{ etc.})$$

thereupon lowering the characteristics of x', y', and z' by the harmless device of subtracting and adding unity to each logarithm, arriving finally at the form

$$x'' = a'' + by'' + cz''$$

where the double primes denote suppressed logarithms, namely,

$$x'' = x' - 1 = \log x - 1$$

$$y'' = y' - 1 = \log y - 1$$

$$z'' = z' - 1 = \log z - 1$$

$$a'' = a' - 1 + b + c = \log a - 1 + b + c$$

By transposing the formula all to one side, we have the acceptable form

$$f = x'' - (a'' + by'' + cz'')$$

The derivatives of f are as follows:

$$f_{x'} = 1, \quad f_{y'} = -b, \quad f_{z'} = -c$$

$$f_{a''} = -1, \quad f_b = -y'', \quad f_c = -z''$$

Then

$$L = 1/W = f_{x'}f_{x'}/w_{x'} + f_{y'}f_{y'}/w_{y'} + f_{z'}f_{z'}/w_{z'}$$

$$= 1/w_{x'} + b^2/w_{y'} + c^2/w_{z'}$$

$$= 0.434^2 \{1/x^2w_x + b^2/y^2w_y + c^2/z^2w_z\}$$

the last step coming from exercise 11 b on page 30.

In this investigation, and in others like it, the S.Es. have all been found proportional to the quantities measured; to be specific

The S.E. of x is 7 percent of x

" " " y is 6 " " y

" " " z is 5 " " z

It follows, then, from Eq. 13 on page 12 that

$$x^2w_x = (x\sigma/\sigma_x)^2 = \sigma^2/.07^2$$

$$y^2w_y = (y\sigma/\sigma_y)^2 = \sigma^2/.06^2$$

$$z^2w_z = (z\sigma/\sigma_z)^2 = \sigma^2/.05^2$$

wherefore

$$1/W = (0.434/\sigma)^2\{(.07)^2 + (.06\ b)^2 + (.05\ c)^2\}$$

which is constant throughout, independent of x, y, and z. This is the second of the two important features described above.

Now σ is open to arbitrary choice, since weights are not absolute but are relative only (page 12); and a convenient choice is to put

$$\sigma^2 = 0.434^2 \{.07^2 + (.06\,b)^2 + (.05\,c)^2\}$$

whereupon W <u>becomes unity at all points</u>. The value of σ, in this problem, is not needed until at the end, when it will be compared with σ(ext) (see section 6d, p. 21); what is more important at present,

b and c will not be needed for the calculation of W, in spite of the fact that x, y, and z are all subject to error. From a computational standpoint, this is a fortunate situation, resting on the peculiar combination of the form of the fitted function and the S.Es. of x, y, and z.

As it happens, W being constant (unity) throughout, the same results for a, b, and c would come from normal equations set up under the (incorrect) assumption that only the measurements on volume are subject to error and all of unit weight. But estimates of the parameters, however important, are not the whole problem; one ought also to consider the adjustment of the observations for a study of the trends (if any) in the residuals; one ought also to know the minimized ϕ^2 for considerations of the fit of the formula, as for example, by comparing σ(ext) with the prior σ, which fortunately is at hand in this example as it was also in the preceding one. If the errors in the diameter and merchantable height are masked, none of the residuals in volume, merchantable height, or diameter, can be found; moreover, the entry in Eq. IV of the solution, which should be ϕ^2, is instead an unknown multiple of it, wherefore the possibility of reconciling the known experimental conditions with the fit of the curve is lost or put on a basis that is apt to do more harm than good.

Approximate values of a, b, and c (hence also of a") having been found by some method or other (see page 34), or known from previous experience, one would next calculate the value of

$$f_o = X'' - (a_o'' + b_oY'' + c_oZ'')$$

at each of the 66 points, the capitals referring to the observed values of log x - 1, log y - 1, log z - 1.

Since W = 1 throughout, tables 1 and 2 of section 16 coalesce, and the normal equations are symbolized as follows:

No.	va"	vb	vc	=	1	Sum
I	66	[Y"]	[Z"]		$-[f_o]$	-
2		[Y"Y"]	[Y"Z"]		$-[Y''f_o]$	-
3			[Z"Z"]		$-[Z''f_o]$	-
4					$[f_of_o]$	-

Since most of the adjustment is already contained in the approximate values of a, b, and c, a maximum of two figures would suffice in any column of X", Y", Z", or f_o; and a maximum of three figures would likely suffice in the normal equations. Such a simplification is our compensation for the trouble of computing f_o at each point.

The solution would proceed as on page 107. Eq. IV will contain the minimized ϕ^2, correctly distributed among the residuals in volume, height, and diameter. The reciprocal matrix found in Eqs. 11, 12, and 13 will contain the variance and product variance coefficients of a", b, and c.

<u>Exercise.</u> Express the variance coefficients of a' and a in terms of c_{11}, c_{12}, etc. found in the reciprocal matrix for a", b, and c.

Instead of using approximate values of a, b, and c, and computing an f_o at each point, Mr. Buell had already adopted the somewhat longer process of using $a_o' = b_o = c_o = 0$, $f_o = X'$, and solving for a, b, and c directly. His normal equations are symbolized as follows, directly in terms of the logarithms on page 170.

No.	a	b	c	=	1	Sum
I	66	[Y']	[Z']		[X']	-
2		[Y'Y']	[Y'Z']		[Y'X']	-
3			[Z'Z']		[Z'X']	-
4					[X'X']	-

Numerically, his equations are these:

No.	a	b	c	=	1
I	66	102.451000	84.090000		152.136000
2		159.921325	131.022337		237.958322
3			107.853544		195.795651
4					356.809522

The solution was found to be:

$$a' = 1.78222, \quad a = 0.01652$$

$$b = 0.87476$$

$$c = 2.14226$$

Hence the relation found is

$$x = 0.0165\ y^{.875} z^{2.14}$$

Not having at hand the complete form of solution, and in particular, not having ϕ^2 as it would appear in the form of solution shown in section 17 page 107, we shall here make use of exercise 3 of section 17 (see also exercise 25 of section 19), thus getting

$$\begin{aligned} \phi^2 = 356.809522 &+ 152.136000 \cdot 1.78222 \\ &- 237.985322 \cdot 0.87476 \\ &- 195.795651 \cdot 2.14226 \\ = 0.324 & \end{aligned}$$

It will be noted that ϕ^2 is here the small remainder left over from the addition and subtraction of relatively much larger numbers. To secure two figures in ϕ^2, one must carry a, b, and c through the fourth decimal; this is so in spite of the fact that we can not possibly rely statistically on so many figures in a, b, and c, a fact that would be evident from their S.Es. or from forest measurements in general. This situation is to be contrasted with the relatively few figures that would be required for the normal equations if good approximate values a_o'', b_o, and c_o had been used for the calculation of f_o at every point; with good approximations, the sum $[f_o f_o]$ would itself be close to the minimized ϕ^2, so that the correction terms need not be carried far. The reader will realize that this matter has been stressed earlier; see e.g. pages 89, 99, 119, 120, and 124.

We can now make the external estimate of σ from the value of ϕ^2 computed above, using Eq. 21 page 19 with the result that

$$\sigma^2(\text{ext}) = \phi^2/(66-3) = 0.324/63 = 0.00514$$

This is to be compared with the prior σ^2, which from the choice made in terms of b and c on page 171 turns out to be

$$\begin{aligned} \sigma^2 &= 0.434^2 \{.07^2 + (.06 \cdot .875)^2 + (.05 \cdot 2.142)^2\} \\ &= 0.00356 \end{aligned}$$

Thus $\sigma^2(\text{ext})$ is about 50 percent larger than the prior σ^2. The possibility of this comparison is the third feature mentioned at the start.

A more exact comparison of the two estimates of σ can be made as follows. First of all, we need χ^2. By Eq. 1a on page 5,

$$\chi^2 = \phi^2/\sigma^2 = 0.324/0.00356 = 91.0$$

Since tables of chi-square do not run so high as 63 degrees of freedom we use Fisher's function*

$$\sqrt{2\chi^2} - \sqrt{2k - 1}$$

which works out to be 2.3, giving P = about 0.02. This is really not a bad value of P for an empirical formula of this kind, though it might be well to look carefully at the data for inhomogeneities of various kinds.

It would be interesting to make a study of the residuals as functions of x, or y, or z, but we shall not stop here except to indicate how the residuals would be computed. From Eqs. 87 page 86 we have

$$\lambda = X'' - (a'' + b\ Y'' + c\ Z'') \qquad \text{at point } h$$

whence by Eqs. 89 on page 91 the logarithmic residuals are

$$\left.\begin{aligned} V_{x'} &= (1/w_{x'})\lambda\ f_{x'} = \lambda/w_{x'} = (0.434 \cdot 0.07/\sigma)^2 \lambda \\ V_{y'} &= (1/w_{y'})\lambda\ f_{y'} = -\lambda\ b/w_{y'} = -(0.434 \cdot 0.06/\sigma)^2 \lambda\ b \\ V_{z'} &= (1/w_{z'})\lambda\ f_{z'} = -\lambda\ c/w_{z'} = -(0.434 \cdot 0.05/\sigma)^2 \lambda\ c \end{aligned}\right\} \quad \text{at point } h$$

since $df/dx' = df/dx''$, etc.

The special features peculiar to this problem have been covered, and the remaining details will be omitted; the reader, however, will profit from Professor Schumacher's comments on the foregoing, which commence over the page.

---

* This function is written at the bottom of table III in Fisher's Statistical Methods for Research Workers (Oliver and Boyd), all editions. When k is large, say above 30, it is distributed very nearly as a normal deviate with unit S. D.

Comments from Professor Francis X. Schumacher, Duke University

1. The number of figures required in the solution of Mr. Buell's normal equations could be cut down by the calculation of an f_o at every point, as you have emphasized, but perhaps a more effectual saving of labor would follow upon transferring the origins of coordinates from the natural zeros to the logarithmic means $\overline{X'}$, $\overline{Y'}$, $\overline{Z'}$. We know from the first normal equation of either the set at the bottom of page 172 or that in the middle of page 173 that the fitted plane will pass through the logarithmic means, i. e. the final values of a, b, and c will satisfy

$$X' = a' + b\overline{Y'} + c\overline{Z'}$$

The transfer of the origins will not only cut down on the number of figures required, but will also eliminate the parameter a' and reduce the number of normal equations by one, leaving only b and c as the unknowns, a' to be found afterward by noting that

$$a' = \overline{X'} - b\overline{Y'} - c\overline{Z'}$$

The new sums and cross products (to be denoted by appending ° to the brackets) would be found by making the following reductions from Mr. Buell's equations:

$$
\begin{aligned}
[Y'Y']^\circ &= 159.921325 - 102.451^2/66 &&= 0.887880\\
[Y'Z']^\circ &= 131.022337 - 102.451 \cdot 84.090/66 &&= 0.490450\\
[Z'Z']^\circ &= 107.853544 - 84.090^2/66 &&= 0.715240\\
[Y'X']^\circ &= 237.985322 - 102.451 \cdot 152.136/66 &&= 1.826453\\
[Z'X']^\circ &= 195.795651 - 84.090 \cdot 152.136/66 &&= 1.960557\\
[X'X']^\circ &= 356.809522 - 152.136^2/66 &&= 6.122213
\end{aligned}
$$

Four decimals will suffice, whereupon Mr. Buell's normal equations (p. 173) reduce to the following set, which can be solved as shown.

No.	b	c	=	1	Sum
I	0.8879	0.4904		1.8265	3.2048
2		.7152		1.9606	3.1662
3	Factors			6.1222	9.9073
4	-0.55231	-.2708		-1.0088	-1.7700
II		.4444		0.9518	1.3962✓
5	-2.05710			-3.7573	-6.5926
6	-2.14176			-2.0385	-2.9903
III				0.3264	0.3264✓
8		b	=	0.8742	
7		c	=	2.1418	3.1418✓

The values of b and c just obtained agree well enough with those on page 173, but with fewer figures and less trouble; and the same can be said for the sum of squares 0.3264 seen in Eq. III. Otherwise obtained,

$$\phi^2 = 6.1222 - 0.8742 \cdot 1.8265 - 2.1418 \cdot 1.9606$$

$$= 0.3263$$

affording an interesting check. (The two figures 0.3264 and 0.3263 for ϕ^2 show a numerical comparison of the two expressions in parts c and a respectively of exercise 3 on page 109).

2. I should like to offer the following in the hope of fostering first approximations as a preliminary to the real work of fitting by least squares. If the merchantable portion of the tree stem were of the same geometrical form in all tree sizes, the volume would vary directly with the height and as the square of the diameter. Hence useful approximations should be

$$b = 1$$

$$c = 2$$

$$a' = \overline{X'} - \overline{Y'} - 2\overline{Z'} \qquad \text{(not needed in the plan just outlined)}$$

The problem is then really one of finding the effect of <u>changes</u> produced by the form of the merchantable solid upon tree volume.

INDEX

INDEX

INDEX